



Research Article

Received: date: 20.03.2024 Accepted: date: 17.05.2024 Published: date: 30.06.2024

turkeyelections: The Most Comprehensive Initial R Package Developed on Election Results in Turkey

Ozancan Özdemir¹

¹Department of Statistics, Middle East Technical University; ozancan@metu.tr Orcid: https://orcid.org/0000-0002-7850-3885¹ *Correspondence: ozancan@metu.edu.tr

Abstract: Election results offer more than just the vote percentages of the participating parties and candidates; they also enable valuable inferences across various domains, particularly about the characteristics of the society in which the elections are conducted, thereby attributing significance to the data on election outcomes. However, within the context of Turkey, it is observed that datasets containing election results in an easily accessible and processable format are limited. Of the 45 elections held from 1946 to 2023, only data for 13 are shared by the Supreme Election Council responsible for the elections. To address this issue in such a popular field, the user-friendly turkeyelections package has been developed within the R programming language environment. This package facilitates easy access and utilization of the results of 16 elections conducted in Turkey between 1995 and 2023, on both national and provincial levels, and has been available for use since February 2024. This study presents details and usage information about the turkeyelections package, the most comprehensive database on Turkish election results developed to date. It enables easy downloading, processing, merging with other datasets, and visualization within the R environment, marking it as the first R programming language package developed for election results in Turkey.

Keywords: R programming; Turkey electoral data; Data Visualization; Open Data

1. Introduction

In the political sphere, an election is characterized as a procedure through which individuals or bodies tasked with public responsibilities are selected by a designated electorate from a pool of multiple candidates [1]. Since the shift to a multi-party framework in 1946, Turkey has conducted a total of 45 electoral events up until the second stage of the presidential elections on May 28, 2023. This total comprises 21 national or parliamentary elections, 13 local or municipal governance elections, 7 referenda, and 4 presidential elections [2]. It should be noted that the actual number of local elections held from 1946 to 2023 exceeds the 12 initially cited. However, statistical counting of local elections began with those held on November 17, 1963, following the implementation of Law No. 307 in 1963, which instituted the direct election of mayors via a simple majority system. Furthermore, Turkey's inaugural referendum occurred on July 9, 1961, while the first presidential elections permitting direct public voting were held on August 10, 2014 [2].

The 1965 parliamentary elections in Turkey employed the national remainder method for seat allocation. Subsequent elections in 1969, 1973, and 1977 favored the d'Hondt system, which was implemented without an electoral threshold. Following the 1980 military coup, significant alterations to the electoral framework were mandated by the 1982 constitution. This legal reform transitioned from the previous threshold-free d'Hondt method to a proportional representation system that included a national threshold of 10% and additional district-level thresholds. Moreover, the 1982 constitution led to the dissolution of the bicameral legislature, which consisted of the National Assembly and the Senate

Citation: O. Özdemir, " turkeyelections: The Most Comprehensive Initial R Package Developed on Election Results in Turkey ", Journal of Statistics and Applied Sciences, no. 9, pp. 67-76, June. 2024, doi:10.52693/ jsas.1456233

of the Republic, both established by the 1961 constitution. Despite some subsequent modifications, the core principles of the electoral system have remained largely unchanged [3].

Election outcomes provide more than just the identification of winners; they also offer valuable insights into the societal context when analyzed alongside data from various domains such as economics, sociology, or social media sentiments. For instance, Akarca and Tansel explored the influence of governmental economic performance on voter behavior in Turkey by examining parliamentary and local election results from 1950 to 2004 [4]. Erener's research assessed the relationship between the consumer confidence index and the electoral performance of political parties CHP and AKP during the elections held between 2004 and 2019 [5]. Additionally, Doğan and Tokgöz investigated the impact of microeconomic factors on electoral decisions by analyzing the results of 11 parliamentary elections in Turkey from 1983 to 2018. In their study, they developed a comprehensive dataset by amalgamating information from various sources [6].

The scrutiny of election outcomes extends beyond the economic sphere, incorporating geographical analyses to discern spatial patterns in voting behavior. The 2011 parliamentary elections in Turkey, for instance, were examined from a geographical perspective to determine whether the results mirrored any specific spatial structures [3]. Gündem's research further integrated spatial statistics with electoral data by analyzing the vote percentages of political parties across provinces in the 2015 elections, utilizing a least-squared regression analysis to explore the influence of geographic location on voting patterns [7]. Similarly, Yiğit and Sezgin investigated the interplay between politics and space by reviewing general election results from 2000 to 2015, aiming to understand spatial variations in political preferences [8]. In a more focused study, Gülhan applied both OLS regression and spatial lag regression techniques to analyze the 2019 local election results in Istanbul, seeking to uncover underlying shifts in the city's political landscape [15].

The aforementioned studies highlight the critical role that electoral data analysis plays in interpreting various societal dynamics, including behavioral patterns. The availability and frequency of such studies are often contingent upon the accessibility of data. For example, in the United States, there are numerous open data platforms specifically dedicated to electoral information, facilitating extensive research. Within the context of R programming, tools like the 'election' package provide comprehensive datasets on U.S. presidential election results dating back to 1920 [16]. Similarly, the 'electionsBR' package, designed for managing Brazilian electoral data, has been utilized extensively since its introduction in 2016, with over 40,000 downloads [17].

Conversely, conducting similar research in Turkey presents significant challenges primarily due to data accessibility issues. The Supreme Election Council of Turkey only provides open data formats for 13 out of 45 election results, all post-2011. Researchers are thus compelled to manually process PDF files from the Turkish Statistical Institute or seek alternate sources to access and analyze earlier election results. The lack of a comprehensive and easily accessible electoral database in Turkey significantly impedes the ability to efficiently utilize electoral data in academic research. This limitation undoubtedly restricts the scope and ease of conducting electoral studies within the Turkish context.

To fill this gap in the literature and create a database that researchers can easily access and process, turkeyelections package was developed in R programming language, which is one of the most popular programming languages in social sciences, and published on the Comprehensive R Archive Network (CRAN) on February 4, 2024. This user-friendly R package is the first R programming package that provides an easy way of downloading and analyzing electoral data in Turkey. The main goal of the study is to introduce this package and give a detailed guide regarding its usage.

The content of this article consists of four parts with this introduction part. The next chapter explains the purpose of the package and gives some details about it. We exemplify the usage of this package in the third chapter. In the last chapter, the overview of this study is summarized, and the gained results from the experiments are summarized. The article is concluded by a discussion about the improvements to be made as a future direction.

2. Purpose and Content of the Package

The turkeyelections package, a user-friendly R package, is the first R package developed to facilitate access to processable data for research processes related to election results in Turkey, catering not only to researchers working in social sciences, particularly in political science and public administration, but also to individuals and institutions in the political research sector, including political parties. The package provides the results of general, local, and presidential elections held in Turkey from the 1995 general elections up to the 2023 general elections, over a span of 18 years, as a processable vector type, namely data.frame and tibble objects, in the R programming language environment. These results are available by vote count and percentage, both nationwide and by province. Additionally, the package facilitates easy export of this data to various databases such as Excel, through commands available in the R programming language, like write.csv. Lastly, the package includes functions that automatically generate graphs such as bar charts and parliament diagrams, making it easier to reach conclusions from the data. The distribution of the elections is shown in the table below.

General Parlamentary Election	Local Election	Presidential Election
1995	2004	2014
1999	2009	2018
2002	2014	2023 First Round
2007	2019	2023 Second Round
2011		
June 2015		
November 2015		
2018		
2023		

Table 1. The Distribution of Elections in turkeyelections Package

The process of dataset compilation from sources like the Turkish Statistical Institute and the Supreme Election Council involved diverse techniques such as web scraping using the rvest package and extracting data from PDF files with the pdftools package in R [9-10]. Post-collection, data manipulation was carried out using the tidyverse package in R [11]. This phase included several critical steps to ensure data integrity and usability: correcting missing observations, amending errors introduced during the scraping process, revising and creating more descriptive titles, removing potentially contaminating elements like percentage signs, and converting variables mistakenly classified as text to numerical form. These measures were essential to maintain the cleanliness and accuracy of the data, adhering to strict academic standards.

Regarding the visualization aspect, the creation of the package's logo was accomplished using the DALL-E model, an AI developed by OpenAI, named after "Denoising Autoencoder for Learned Language Embeddings." It's a generative model capable of creating images from textual descriptions, understanding natural language to produce visuals that closely match the given descriptions [13]. You can see the logo below.



Figure 1. The turkeyelections package logo created by DALL-E

3. Implementation and Technical Details

As mentioned above, the turkeyelections package provides the results of general, local, and presidential elections held in Turkey between 1995 and 2023, in a format processable by the R programming language, under two different categories: nationwide and by province.

The package was published on CRAN on February 4, 2024. It requires a minimum of R version 4.0 to function. It has been developed with the assistance of other R packages, which are also listed in the imports section. To install the turkeyelections package into the R library, these packages must also be installed. In order, these packages are ggplot2, ggpol, stringr, dplyr, colorRamp2, and stats. Detailed technical information is available in the documentation created during the package's upload process to CRAN [12]. Besides, the GitHub page of the package represents some technical details related to the package. The package contains 38 functions. While four of these functions facilitate the use of the package and the production of results, the remaining 34 are data packages names that return the election results as a tibble object in the R programming environment.

This section will discuss and exemplify the use of the four main functions within the package, which are data_list, show_parliament, show_province_and_party, and show_result_for_turkey.

The data_list() function does not require any arguments and can be directly applied by writing data_list() in the R console. This function returns a data.frame object consisting of 2 columns and 34 rows. The column names, election_names, explain whether the results pertain to the provinces or the entire country in Turkish, while the data_name column shows the R function name that should be used to obtain the election results corresponding to the explanations. The application of this command is shown in the figure below.

> data_list()			
election_names	data_name		
1 2014 Cumhurbaşkanlığı Seçimi (İllere göre sonuçlar)	cb_2014_il		
2 2014 Cumhurbaşkanlığı Seçimi (Türkiye geneli)	cb_2014_turkiye		
3 2018 Cumhurbaşkanlığı Seçimi (İllere göre sonuçlar)	cb_2018_il		
4 2018 Cumhurbaşkanlığı Seçimi (Türkiye geneli)	cb_2018_turkiye		
5 2023 Cumhurbaşkanlığı Seçimi 2.Tur (İllere göre sonuçlar)	cb_2023_ikinci_tur_il		
6 2023 Cumhurbaşkanlığı Seçimi 2.Tur (Türkiye geneli) cb_20	023_ikinci_tur_turkiye		
7 2023 Cumhurbaşkanlığı Seçimi 1.Tur (İllere göre sonuçlar)	cb_2023_ilk_tur_il		
8 2023 Cumhurbaşkanlığı Seçimi 1.Tur (Türkiye geneli) ct	p_2023_ilk_tur_turkiye		
9 1995 Milletvekili Genel Seçimi (İllere göre sonuçlar)	genel_secim_1995_il		
	nel_secim_1995_turkiye		
11 1999 Milletvekili Genel Seçimi (İllere göre sonuçlar)	genel_secim_1999_il		
12 1999 Milletvekili Genel Seçimi (Türkiye geneli) ger	nel_secim_1999_turkiye		
13 2002 Milletvekili Genel Seçimi (İllere göre sonuçlar)	genel_secim_2002_il		
	nel_secim_2002_turkiye		
15 2007 Milletvekili Genel Seçimi (İllere göre sonuçlar)	genel_secim_2007_il		
16 2007 Milletvekili Genel Seçimi (Türkiye geneli) ger	nel_secim_2007_turkiye		
17 2011 Milletvekili Genel Seçimi (İllere göre sonuçlar)	genel_secim_2011_il		
	nel_secim_2011_turkiye		
19 2018 Milletvekili Genel Seçimi (İllere göre sonuçlar)	genel_secim_2018_il		
	nel_secim_2018_turkiye		
21 2023 Milletvekili Genel Seçimi (İllere göre sonuçlar)	genel_secim_2023_il		
22 2023 Milletvekili Genel Seçimi (Türkiye geneli) ger	nel_secim_2023_turkiye		
	_secim_haziran_2015_il		
	m_haziran_2015_turkiye		
25 2015 Kasım Milletvekili Genel Seçimi (İllere göre sonuçlar) gene	el_secim_kasim_2015_il		
	cim_kasim_2015_turkiye		
27 2004 Belediye Başkanlığı Seçimi (Türkiye geneli) yer	rel_secim_2004_turkiye		
28 2009 Belediye Başkanlığı Seçimi (Türkiye geneli) yerel_secim	_2009_belediye_turkiye		
29 2009 Büyükşehir Belediye Başkanlığı Seçimi (Türkiye geneli) yerel_secim_20	009_buyuksehir_turkiye		
30 2009 Belediye Başkanlığı Seçimi (İllere göre sonuçlar)	yerel_secim_2009_il		
31 2014 Belediye Başkanlığı Seçimi (İllere göre sonuçlar)	yerel_secim_2014_il		
	rel_secim_2014_turkiye		
33 2019 Belediye Başkanlığı Seçimi (İllere göre sonuçlar)	yerel_secim_2019_il		
34 2019 Belediye Başkanlığı Seçimi (Türkiye geneli) yer	rel_secim_2019_turkiye		

Figure 2. The output of data_list() function on R console

The show_parliament() command is a function that incorporates the use of the ggplot2 functions and geom_parliament function from ggpol packages [19] and operates exclusively on datasets containing results for general elections across Turkey. Developed as a single-argument command that only requires the name of the dataset, it informs the user if the provided dataset is incompatible with the command by stating that the dataset does not match the command's requirements. There is no need to input the dataset as a character object within this function; merely entering the name of the dataset suffices. Below is a typical application of this function and the graph that results from it: > show_parliament(genel_secim_2023_turkiye)

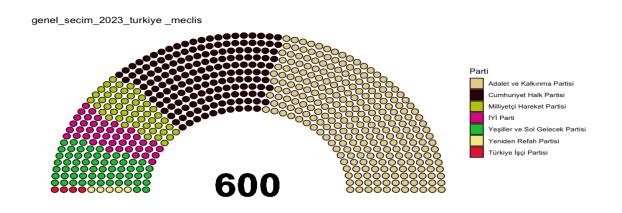


Figure 3. The output by show_parliament() function for the general election results in 2023 in Turkey

As seen above, the parliamentary representation of the parties from the May 2023 election results is visualized using a horseshoe (horseshoe) graph with the help of the show_parliament() function. The name of the dataset written inside the show_parliament() function is automatically displayed as the title. The total number of seats is also shown at the bottom of the plot. Since the resulting graph is a ggplot2 object, further visual manipulations can be applied using ggplot2 package commands such as labs(), geom_text(), allowing for desired modifications on the generated visualization [18]. For instance,

> p1<-turkeyelections::show_parliament(genel_secim_2023_turkiye)</p> p1+ggplot2::labs(title ="2023 Secim Sonuçlarına Göre Meclis Dağılımı")

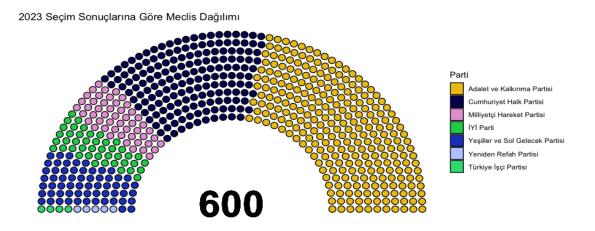
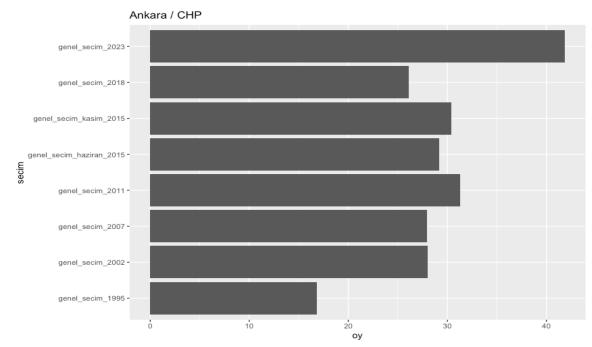


Figure 4. The output by show_parliament() function for the general election results in 2023 in Turkey after manipulation

As understood from these two figures, the function generates a random color for each party each time. To overcome this problem, the users can utilize the scale_fill_manual() function from ggplot2 [18].

The show_province_and_party() command utilizes the election result data in the package to display the vote percentages a party has received in a province over the years. This command has three arguments: province, party, and percentage. The province argument specifies the name of the province

in Turkey to be examined, while the party argument specifies the name of the political party to be examined. The percentage argument is a boolean value that, when true, also displays the party's vote percentages as text on the generated graph. The default value of this argument is FALSE, but it can be changed to TRUE if desired. An example use of this command is demonstrated.



> show_province_and_party("Ankara","CHP",percentage = FALSE)

Figure 5. The output by show_province_and_party() function that shows the vote percentage of the Republican People Party in Ankara over time

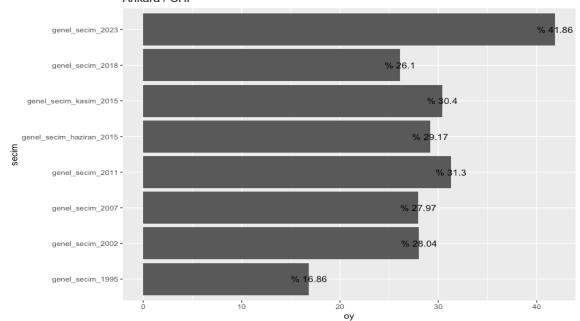


Figure 6. The output by show_province_and_party() function that shows the vote percentage of the Republican People Party in Ankara over time not only in bars but also in text

Just like the previous graph, since the show_province_and_party command returns a ggplot2 object, this object can also be manipulated and its visual appeal enhanced using functions within the ggplot2 package. For example,

> p1<-show_province_and_party("Ankara","CHP")+geom_col(fill="darkred")+labs(title = "CHP'nin Ankara'daki Seçim Sonuçları") > p1

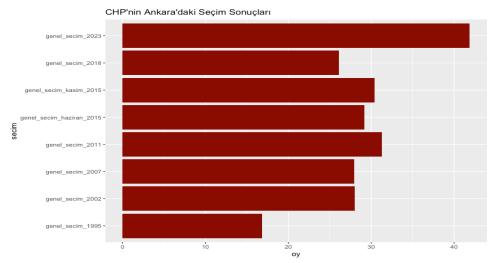


Figure 7. The output by show_province_and_party() function that shows the vote percentage of the Republican People Party in Ankara over time after ggplot2 manipulation

Finally, the show_result_for_turkey() function displays the results of elections across Turkey. This function operates based on two arguments: the dataset and percentage. The dataset argument, like in the previous two functions, represents the name of the dataset to be examined, while the percentage argument, similar to the show_province_and_party() command, serves to display the vote percentages received by parties as text on the generated visual. The default value of this argument is FALSE, but it can be activated with the command percentage=TRUE. The show_result_for_turkey() function can be used for general, local, and presidential election results that have nationwide data. If the dataset argument used within this function does not cover nationwide results, the function will return a warning message indicating that the dataset is incorrect. An example usage is shown below. > show_result_for_turkey(genel_secim_2023_turkiye)

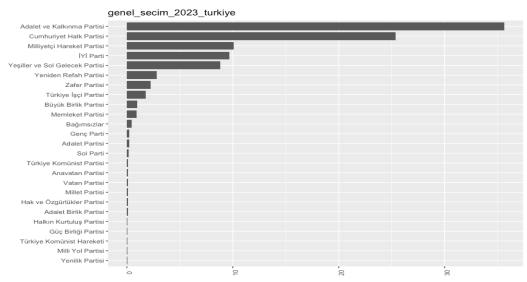
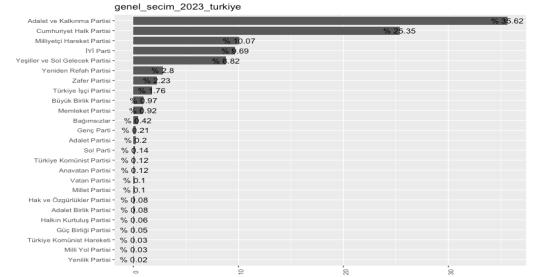


Figure 8. The output by how_result_for_turkey() function for the election results in 2023



> show_result_for_turkey(genel_secim_2023_turkiye,percentage = TRUE)

Figure 9. The output by how_result_for_turkey() function showing the election results in 2023 not only in bars but also in text

Like the other two graphical objects, the ggplot2 object resulting from this command can also be manipulated with relevant commands, enhancing its visual quality.

4. Conclusion and Future Changes

The article introduces the turkeyelections package, the inaugural R package created in Turkey designed to provide access to a comprehensive collection of election results. This package encompasses the outcomes of 16 elections held between 1995 and 2023, which include local, general, and presidential elections, compiled into 34 datasets that cover both provincial and national results.

This package marks the first time such a significant dataset on Turkish election results has been made accessible for analysis within the R environment. Researchers can now visualize and conduct various statistical analyses thanks to the robust capabilities of the ggplot2 framework and its extensions like the ggpol packages. These tools support the creation of complex visual representations of electoral data, facilitating a deeper understanding of the dynamics at play in Turkish elections.

Aimed at enhancing the accessibility and analysis of electoral data, the turkeyelections package provides researchers with powerful commands to streamline their workflows. Since its release on CRAN, the package has been downloaded around 500 times, reflecting its significance and utility within the academic and data analysis communities. This uptake underscores the growing interest and need for sophisticated tools to analyze election data in Turkey, promising to spur further research and insights into electoral trends and patterns

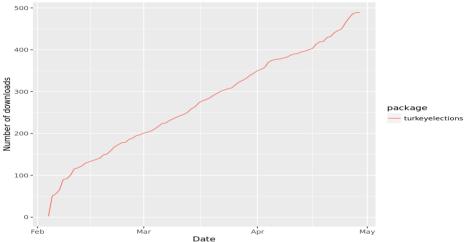


Figure 10. The cumulative number of downloads of turkeyelection package from CRAN [14]

While the turkeyelections package significantly simplifies the processing of Turkey's electoral data in the R environment, there are areas identified for future improvement. One of the primary concerns to be addressed in the subsequent development phases is the inconsistency in party names across datasets. Efforts will be made to ensure uniformity in party identifiers to enhance the reliability and analysis capabilities of the package.

Another major development will be to standardize the format of election results for elections conducted prior to 1995, beginning with the inclusion of results from the upcoming March 31, 2024, election. This step will help to create a more comprehensive and consistent dataset, facilitating longitudinal studies and trend analysis.

The scope of the local elections datasets also requires expansion. Currently, the data obtained from the Supreme Election Council includes vote details only for municipality councils, which limits the depth of analysis possible. A revision and expansion of these datasets to include additional layers of local election results will provide a richer resource for detailed local political analysis.

Furthermore, there are plans to enhance the package's functionality by introducing new features, such as the ability to generate customized maps of Turkey. These maps will highlight electoral results specific to parties and years, complementing the existing graphical outputs provided by the package. This new feature aims to visually represent electoral dynamics across different regions, offering a more intuitive understanding of electoral trends and their geographical distributions.

These improvements will not only refine the existing capabilities of the turkeyelections package but also expand its utility for researchers and analysts studying Turkish elections, thereby contributing to more nuanced insights into the political landscape of Turkey.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] H. S. Türk, "Seçim, Seçim Sistemleri ve Anayasal Tercih," Anayasa Yargısı, vol. 22, no. 1, pp. 75-113, 2006.
- [2] YSK Web Portal, "Tarihçe," 2024. [Online]. Available: https://www.ysk.gov.tr/tr/tarihce/1510.
- [3] Ö. Kaya, E. Toroğlu, and F. Adigüzel, "2011 Genel Seçimlerinde Partilerin Aldığı Oy Oranlarının İlçeler Ölçeğinde Mekânsal Analizi," Coğrafya Dergisi, no. 31, Article 31, 2016.
- [4] A. T. Akarca and A. Tansel, "Economic Performance and Political Outcomes: An Analysis of the Turkish Parliamentary and Local Election Results Between 1950 and 2004," Public Choice, vol. 129, no. 1, pp. 77-105, 2006. [Online]. Available: https://doi.org/10.1007/s11127-005-9013-9.
- [5] C. Erener, "Tüketici Güven Endeksi İle Seçim Sonuçlari Arasındaki İlişki: Ak Parti Ve Chp Üzerine Ekonometrik Bir Analiz," 9 July 2021. [Online]. Available: https://dergipark.org.tr/en/pub/jeaa/issue/64041/939900.
- [6] A. Doğan and E. Tokgöz, "Seçmen Tercihinde Ekonomik Gelişmelerin Rolü: Ekonomik Oy Verme Teorisi Bağlamında Türkiye Genel Seçimleri Üzerine Bir İnceleme," e-gifder, vol. 10, no. 2, pp. 1179-1205, 2022. [Online]. Available: https://doi.org/10.19145/e-gifder.1112826.
- [7] F. Gündem, "Türkiye'de Seçmen Davranışlarının Belirleyenleri: Mekânsal Ekonometrik Bir Yaklaşım," 2017.
- [8] H. H. Yiğit and A. Sezgin, "Türkiye'de Milletvekili Genel Seçimleri Üzerine Bir Mek Nsal Bağimlilik Analizi," Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, no. 35, Article 35, 2019.
- [9] H. Wickham, "rvest: Easily Harvest (Scrape) Web Pages," R package version 1.0.4, 2024. [Online]. Available: https://github.com/tidyverse/rvest, https://rvest.tidyverse.org/.
- [10] J. Ooms, "pdftools: Text Extraction, Rendering and Converting of PDF Documents," R package version 3.4.0. [Online]. Available: https://docs.ropensci.org/pdftools.
- [11] H. Wickham et al., "Welcome to the tidyverse," Journal of Open Source Software, vol. 4, no. 43, Article 1686, 2019. [Online]. Available: https://doi.org/10.21105/joss.01686.
- [12] O. Ozdemir, "turkeyelections: The most comprehensive R package for Turkish Election Results," R package version 0.1.0, 2024. [Online]. Available: https://github.com/ozancanozdemir/turkeyelections.
- [13] M. Fezari, A. Al Dahoud and A. Al-Dahoud, "Augmanting Reality : The Power of Generative AI", 2023. Available: https://www.researchgate.net/publication/370821758_Augmanting_Reality_The_Power_of_Generative_AI
- [14] ipub. 2017. "R Pkg Download Stats Ipub." Ipub. Retrieved (<u>https://ipub.com/dev-corner/apps/r-package-downloads/</u>).
- [15] S. T. Gülhan, "The Election Day that Lasted 84 Days: Mapping the Electoral Geography of the 2019 Istanbul Metropolitan Mayoral Race", 02-Apr-2022. [Online]. Available: osf.io/preprints/socarxiv/ufvtz.
- [16] Wal, Willem M. van der. "Elections: USA Presidential Elections Data." R-Packages, 11 Dec. 2023, cran.r-project.org/web/packages/elections/index.html. Accessed 1 May 2024.

- [17] Silva, Denisson, et al. "ElectionsBR: R Functions to Download and Clean Brazilian Electoral Data." www.academia.edu,www.academia.edu/30307653/electionsBR_R_Functions_to_Download_and_Clean_Brazilian_Elector al_Data. Accessed 1 May 2024.
- [18] Wickham, Hadley, et al. "Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics." R-Packages, 3 Apr. 2023, cran.r-project.org/web/packages/ggplot2/.
- [19] Tiedemann, Frederik. "Ggpol: Visualizing Social Science Data with "Ggplot2."" R-Packages, 8 Nov. 2020, cran.r-project.org/web/packages/ggpol/index.html. Accessed 1 May 2024.