

thstats: An Open-Data R Package for Exploring Turkish Higher Education Statistics

İstatistikler: Türkiye'deki Yükseköğretim İstatistiklerini Keşfetmek için Açık Veri R Paketi

Mustafa Çavuş¹ , Olgun Aydın² 

¹Department of Statistics, Eskişehir Technical University, Eskişehir, Türkiye

²Department of Statistics, Gdansk University of Technology, Gdansk, Poland

Özet

Resmi istatistik, finans, eğitim ve çeşitli diğer alanlar için açık veri kümeleri mevcuttur. Açık veri kümeleri, resmi makamların yanı sıra üçüncü taraflar tarafından yayınlanmaktadır. Örneğin, Türkiye Yükseköğretim Kurumu, Türkiye'deki yükseköğretime yönelik bir web portalı sunmaktadır. Portaldan üniversiteler, fakülteler ve bölümler hakkında detaylı veri setleri elde edilebilmektedir. Portal tarafından sağlanan veriler kullanılarak, yükseköğretimin ekonomi üzerindeki etkisinin anlaşılması, öğrencilerin yükseköğretim açısından zaman içindeki tercihlerindeki değişikliklerin incelenmesi gibi detaylı analizlerin yapılabilmesi mümkündür. Bu portal, birçok önemli bilgil içermesine rağmen, herhangi bir Uygulama Programlama Arayüzü (API) veya verileri tek bir yerden kolayca indirmek veya sorgulamak için herhangi bir seçenek sunmamaktadır. Mevcut makale, yükseköğretim istatistiklerini kolayca erişilebilir kılmak için tasarlanmış kullanıcı dostu bir R veri paketini tanıtmaktadır. Bu paket araştırmacılara, hali hazırda portaldan kazanmış olan verileri sorgulayabilme ve bazı basit analizler yapabile olanağı sunmaktadır. Paket ile araştırmacıların yükseköğretim istatistiklerine kolayca erişebilmesine katkı sunulması hedeflenmektedir.

Anahtar Kelimeler: Türk yükseköğretimi, istatistik, üniversite, açık veri, R, sosyoekonomik

Open data has become an important element for the reproducibility of scientific studies. In addition to the importance of providing open data, making the data easily accessible on the web has been critical. The importance of open data portals for transparency in the public sector is one of the most crucial topics in current debates on how to provide accountable, participatory, and responsive governance (Lnenicka and Nikiforova, 2021). For these reasons, the push to make data easily accessible has gained momentum in recent years.

Abstract

There are open datasets available for official statistics, finance, education, and a variety of other domains. The open datasets are published by third-party vendors as well as official authorities. For example, The Turkish Higher Education Council maintains a web portal dedicated to higher education in Türkiye. Detailed datasets about universities, faculties, and departments can be obtained from the portal. Using the data provided by the portal, detailed analysis can be done such as the understanding impact of higher education on activities in the economy connected to movements of the students within or between cities. Especially such activity can be observed in real estate since students rent flats or rooms. Moreover, this data helps researchers analyze changes in students' preferences in terms of higher education over time. Although this site contains a variety of important information, it does not provide any Application Programming Interfaces (API) or any other possibilities to easily download or query the data in one place. This paper introduces thstats, a user-friendly R data package designed to make higher education statistics easily accessible. Researchers can use the package to query data, which is already scraped from the portal, using the R functions provided by the package. Thanks to the package, researchers do not need to perform any further effort to delve into Turkish higher education statistics. It is sufficient for the researchers to use the package to begin exploring.

Keywords: Turkish higher education, statistics, university, open-data, R, socioeconomics

The initiatives such as *rOpenSci*¹ and *rOpenGov*² focusing on this issue have provided significant contributions in recent years. *rOpenSci* is a non-profit organization that develops community-contributed R software tools that help to reach web-based scientific data sources efficiently (Boettiger et al., 2015). *rOpenGov* is a community of R package developers dedicated to open government data and related subjects.

1 <https://ropensci.org/>

2 <https://ropengov.org/>

İletişim / Correspondence:

Dr. Mustafa Çavuş
Eskişehir University, Faculty of Science,
Tepebaşı, Eskişehir
e-posta: mustafacavus@eskisehir.edu.tr

Yükseköğretim Dergisi / TÜBA Higher Education Research/Review (TÜBA-HER), 13 (1), 1-7. © 2023 TÜBA
Geliş tarihi / Received: Ağustos / August 28, 2022; Kabul tarihi / Accepted: Şubat / February 22, 2023
Bu makalenin atıf künyesi / How to cite this article: Çavuş, M. & Aydın, O. (2023). Thstats: An Open-Data R Package for Exploring Turkish Higher Education Statistics. *Yükseköğretim Dergisi*, 13(1), 1-7. doi: 10.2399/jod.23.1167895

Bu çalışma, 24-26 Kasım 2021 tarihlerinde Romanya Ulusal İstatistik Enstitüsü tarafından düzenlenen The Use of R in Official Statistics 2021 konferansında sözlü bildiri olarak sunulmuştur. (This study was presented as an oral presentation at The Use of R in Official Statistics Conference 2021 which is organized by Romanian National Statistics Institute on November 24-26, 2021.)

ORCID: M. Çavuş: 0000-0002-6172-5449; O. Aydın: 0000-0002-7090-0931

The *eurostat* is one of the packages was developed by *rOpenGov* (Lahti et al., 2017). The package is designed to search, download, manipulate, and visualize the open data from Eurostat³. There are also several R packages, developed by individual researchers, which provide the ability to analyze open data from various sources in R (Senyondo et al., 2021; Glogowski et al., 2019). While some of the packages scrape web pages that provide open-source data, some of them are using an API. (Abbott, 2019; Freire and McDonnell, 2018; Odell, 2020). Some papers provide built-in R commands to access the open data via APIs (Koontz et al., 2020; McGuinness et al., 2020; Sparks, 2018). The number of open data sources for different scientific disciplines has been increasing with the help of rapidly developing open-source technologies. Thanks to this, researchers have a chance to access various open data sources and this allows researchers to have a different perspective and deep understanding of the topics they focus on. Undoubtedly, open data sources for the higher education systems are also very crucial, as higher education affects the economy and development of the country.

The higher education system in Türkiye is administrated by the Council of Higher Education (CoHE) which is an autonomous institution that is responsible for the planning, coordination, and governance of higher education in Türkiye under the Turkish Constitution and the Higher Education Laws⁴. Turkish higher education consists of three cycles: bachelor's, master and doctoral studies. This paper focuses on the data for admission to the first cycle, which involves 4 to 8 semesters or more, depending on the field of study (Turkish Council of Higher Education, 2021).

The Turkish higher education admission procedure is nationally standardized, and universities may use additional criteria for a further selection of applicants. A national two-stage placement system regulates entry to tertiary education and places students in various programs. The Basic Proficiency Test (Temel Yeterlilik Testi - TYT), formerly known as the Transition to Tertiary Education Examination (Yükseköğretime Geçiş Sınavı - YGS), is a multiple-choice evaluation of key disciplines such as Turkish grammar, social sciences, mathematics, and science. Passing The Basic Proficiency Test is enough to gain admission to short-cycle tertiary programs in which the majority of students enroll. Students must take an additional test called the Field Qualification Test (Alan Yeterlilik Testi - AYT), previously known as the Undergraduate Placement Exam (Lisans Yerleştirme Sınavı - LYS), in subjects relevant to their preferred field of study to enroll in bachelor's programs. Students' preferences, The Basic Proficiency Test and Field Qualification Test results, and average classroom grades in upper secondary education are utilized to decide their placement in bachelor's programs using a centralized system that automatically allocates applicants to study programs. Most programs have a demand-based admissions criterion, while the most selective programs, such as medical, law, and engineering, have a minimum pre-set admissions threshold.

Türkiye does not have a national examination to confirm upper secondary education completion. Instead, upper secondary certification (i.e., receiving a high school diploma) is based primarily on students' average scores on classroom tests from grades 9 to 12⁵.

There are many research studies in the literature focusing on the higher education system and statistics in the world. One of the most common research questions on this topic is "How to analyze higher education entrance exams?". Nagy and Molontay (2021) argued that it is critical to find a suitable admission procedure that can distinguish between students with high academic potential and those who are likely to become future dropouts. Ferrao and Almeida (2019) argued that the relationship between university entrance exam scores and the academic performance of students after their first-year studies vary depending on various socioeconomic factors. Another critical research question is "How researchers can understand the reasons behind the university preferences of students?". Heiskala et al. (2021) found out that, in Finland, higher education attracts successful students from lower socioeconomic backgrounds. However, successful students from higher socioeconomic backgrounds have a substantially higher probability of enrolling in university studies. Tilak (2020) indicated that professions in electronic engineering, computer science, and information technology have changed so rapidly in recent years that it has caused an unbalanced growth of disciplines, creating imbalances in the labor market in countries such as China, Brazil, and Russia.

As can be seen, there are many studies on higher education. However, the importance of these studies is more critical for Türkiye. From 1971 to 2016, the number of students enrolled in Turkish higher education increased by 700% (Ergen and Cakioglu, 2018). The total number of students in higher education is 8.3 million students: 3.85 million of them are face-to-face students, and the rest are in open education (Yurdakul & Şahin-Demir, 2022). Also, Türkiye has the highest number of students in higher education in Europe, as there are two times more bachelor students than Germany which is the most crowded country in European Union by 2018 (Aydın and Cavus, 2021). The reason behind this jump in some students is the Turkish government's strategy which is aiming to have at least one university in every city (Arap, 2010). However, a recent research study showed that not all of the students registered for university continue their studies (Yurdakul and Şahin-Demir, 2022). This situation has been causing many changes in Türkiye and has been having a vital effect on the economy, and socio-demography of the country. These changes can be concluded that higher education in Türkiye has many aspects that need to be analyzed from different perspectives. Several studies are focusing on changes and aspects of Turkish higher education.

³ <https://ec.europa.eu/eurostat>

⁴ <https://www.yok.gov.tr/en/institutional/higher-education-system>

⁵ <https://www.yok.gov.tr/en/institutional/higher-education-system/Id=/content/component/71ee93b4-en>



Apaydin (2020) indicated that the majority of young people prefer to study for a job in the health, engineering, and teaching domains, as those domains offer relatively better employment prospects in Türkiye. Suna et al. (2020) examined the socioeconomic background and preferences of high school students focusing on science regarding the transition into higher education. The study showed that students from higher socioeconomic backgrounds have a higher tendency to begin university studies compared to students from lower socioeconomic backgrounds. Karšli and Anlı (2010) pointed out that the higher education entrance exam is a source of anxiety among young people in Türkiye.

In addition to these studies, a few studies in the literature focus on the rapid development of the higher education system in Türkiye (Tekneci, 2016; Gok, 2016; Ozoglu et al., 2016). There are many studies in the literature about Higher Education in Türkiye. However, due to the limitation of accessing open data sources, the topic could not be analyzed from different perspectives. At this point, the only data source is a web portal⁶ maintained by the Council of Higher Education dedicated to higher education statistics in Türkiye. Detailed statistics about universities, faculties, and departments such as the number of students placed, their scores, gender distribution, number of choices, etc.⁷, and also the success of the graduates of the programs in the central specialization exams can be obtained from the portal. However, there is no easy way to download or query the data using one of the most preferred languages among researchers such as R. This undoubtedly makes it difficult to use the data in research and compels researchers to work on samples. The main goal of the study is to introduce the user-friendly R data package *thestats* which intended to make higher education statistics provided by the Turkish Higher Education Council accessible, except the statistics related to the success of the graduates. Researchers can use the package to easily explore the data, which is already scraped from the portal. The package is expected to be an important source for solving the problem of data-driven planning and policy development (Erdogan, 2014; Günay and Özer, 2016; Ozoglu et al., 2016), which is known as one of the most important problems in the Turkish Higher Education system.

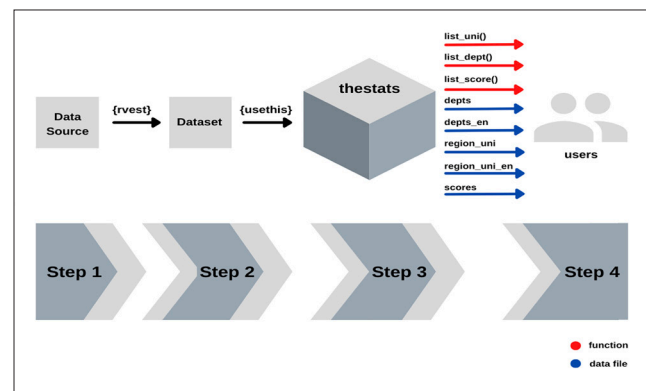


■ Figure 1. Hex logo of thestats package.

This article is organized as follows. The following section provides an overview of the Turkish higher education system. The third section covers information about the development stages of *thestats*, types of data provided by the package, and how the data are obtained. Section four provides information about the usage of the package. The fifth section presents a few potential outcomes that can be obtained by using *thestats*, while the last section summarizes the paper by discussing the importance of the open official data and suggesting potential future studies.

Mechanism of thestats Package

A user-friendly R data package, *thestats*, is developed to make higher education statistics accessible in an easy way. The data provided by this package is gathered from the web portal of the project titled Higher Education Program Atlas⁸.



■ Figure 2. Mechanism of thestats package.

The researchers, who are interested in the data on the portal, have to follow many steps to reach only a part of this very unique dataset. As a first step, the portal requires you to choose a university name or department name. Following this, thirty hamburger menus appear and each of them contains multiple pieces of information. Researchers have to open the menus to see statistics provided through tables or graphs. There is no possibility of downloading the statistics, the only way of retrieving the data is to copy-paste. Considering this complex process, the authors of the package spent a significant amount of time scraping data from the portal carefully. *rvest* package was used to scrape data and proper data manipulation techniques were applied using *dplyr* (Wickham, 2021; Wickham et al, 2021). After ensuring the data quality, R data files are created using *usethis* (Wickham and Bryan, 2021), which is an R package for handling the creation of R data files and wrapping them into an R package. This flow is summarized in ■ Figure 2. The users can simply install *thestats* and begin exploring the Turkish Higher Education Statistics using three functions or direct data files provided by the package.

⁶ <https://yokatlas.yok.gov.tr/lisans-anasayfa.php>

⁷ Full list of the statistics are available on <https://github.com/analyticsresearchlab/thestats/blob/main/docs/Table3.md>

⁸ <https://yokatlas.yok.gov.tr/lisans-anasayfa.php>

■ **Table 1.** Arguments used in the functions of thestats.

Argument	Definition	Function		
		list_dept	list_score	list_uni
region_names	geographical region of universities	✓	✓	✓
city_names	the city where universities are located	✓	✓	✓
university_names	name of universities	✓	✓	
departmen_names	names of departments in universities	✓	✓	
lang	language of the information	✓	✓	✓
aggregation	an option to aggregate the filtered data	✓		✓
var_ids	statistics about the universities in -scores-		✓	

thestats provides three easy-to-use functions: `list_dept()`, `list_uni()`, and `list_score()` with arguments shown in ■ Table 1. The functions simply help the users query and create simple aggregations on the top of data files provided by the package: `depts`, `depts_en`, `regions_cities`, `regions_cities_en`, and `scores`. The arguments provide different options to the users. These options for the `departments_names` are shown in manuals of `depts`, `depts_en` datasets, for the `region_names` and `city_names` arguments are shown in manuals `regions_cities` and `regions_cities_en` datasets. For more information, please refer to the package manual⁹. The default value of the `lang` argument is “en” which defines the language of the results as English, to retrieve results in Turkish, “tr” should be passed to the argument. The aggregate provides two options for `list_uni()` and one option for the `list_dept()` function. These options are `count_by_city` for obtaining the group sum of universities per city, and `count_by_region` for obtaining the group sum of universities per each region defined by the user, respectively. The key argument in the package is the `var_ids`. It is used to call the 196 statistics about the departments, and detailed information about these statistics is given in the manual of `scores` dataset. The various usage examples of the functions and arguments introduced in this section are given in the following section.

Usage of thestats Package

As mentioned in the previous section, *thestats* provides three functions: `list_uni()`, `list_dept()`, and `list_score()`. The `list_uni()` function helps to query universities in cities or regions specified by the user. Moreover, the function allows aggregations such as the number of universities per year in cities or regions specified by the user. The function has four arguments: `region_names`, `city_names`, `aggregation`, and `lang`.

⁹ <https://CRAN.R-project.org/package=thestats>

```
# installing and loading 'thestats' library
install.packages("thestats")
library(thestats)

# listing the universities in Izmir
list_uni(region_names = "all", city_names = "Izmir")

  id      university      type city region      website
1 1031   Dokuz Eylul University State Izmir Aegean http://www.deu.edu.tr
2 1034         Ege University State Izmir Aegean https://ege.edu.tr
3 1120   Izmir Bakircay University State Izmir Aegean https://bakircay.edu.tr
4 1111   Izmir Demokrasî University State Izmir Aegean https://www.idu.edu.tr
5 1057   Izmir Katip Celebi University State Izmir Aegean https://www.ikcu.edu.tr
6 1058   Izmir Institute of Technology State Izmir Aegean https://iyte.edu.tr
7 2035   Izmir University of Economics Private Izmir Aegean https://www.ieu.edu.tr
8 2096   Izmir Tinaztepe University Private Izmir Aegean https://www.tinaztepe.edu.tr
9 2060         Yasar University Private Izmir Aegean https://www.yasar.edu.tr
```

■ **Figure 3.** Example usage of `list_uni()` for listing universities in a specified city.

The `region_names` argument is to pass names of the regions, and `city_names` are for specifying city names shown in the table¹⁰. The `lang` argument is to select English or Turkish as a language for results returned by the function. As an example, let us assume that the user would like to query universities in Izmir, which is a city in the western part of Türkiye in ■ Figure 3.

```
# installing and loading 'thestats' library
install.packages("thestats")
library(thestats)

# listing the number of universities in Izmir and Mugla
list_uni(region_names = "all", city_names = c("Izmir", "Mugla"), aggregation = "count_by_city")

  city type n
1 Izmir Private 3
2 Izmir State 6
3 Mugla State 1
```

■ **Figure 4.** Example usage of `list_uni()` function for listing the number of universities in the specified cities.

In the example, the `list_uni()` function returns information such as the name, type, and web page URLs of the universities in Izmir. The codes in ■ Figure 4 can be used to get the number of universities per type (state or private) in Izmir and Mugla.

¹⁰ <https://github.com/analyticsresearchlab/thestats/blob/main/docs/Table2.md>



The `list_dept()` function helps to query universities and departments by cities or regions. Moreover, it allows aggregations such as the number of universities or the number of universities having specific departments per year in cities or regions defined by the user. The function has a `dept_names` argument in addition to arguments that `list_uni()` has. It allows users to query departments or universities per city or region. As shown in ■ Figure 5, there is a possibility to retrieve universities that have Statistics departments in Izmir and Mugla, using the function.

```
# installing and loading 'thestats' library
install.packages("thestats")
library(thestats)

# listing the number of assistant professors and the incoming exchange students of the Statistics
# departments in Izmir
list_score(region_names = "all",
           city_names = "Izmir",
           university_names = "all",
           department_names = "Statistics",
           var_ids = c("X198", "X196"))
```

id	year	type	program_code	university	faculty	department	X198	X196
1	1831	2820	State	183110239	Dokuz Eylul University Faculty of Science	Statistics	5	0
2	1831	2819	State	183110239	Dokuz Eylul University Faculty of Science	Statistics	6	0
3	1831	2818	State	183110239	Dokuz Eylul University Faculty of Science	Statistics	7	0
4	1834	2820	State	183410227	Ege University Faculty of Science	Statistics	5	0
5	1834	2819	State	183410227	Ege University Faculty of Science	Statistics	5	0
6	1834	2818	State	183410227	Ege University Faculty of Science	Statistics	4	0

■ Figure 5. Example usage of `list_dept()` function for listing universities in the specific cities that have a specified department.

The `list_score()` function is for querying detailed statistics about universities at the level of departments, cities, and regions. The function has `var_ids` in addition to the arguments that `list_uni()` and `list_dept()` have. The `var_ids` allow users to specify the type of statistics described in the table¹¹ and the manual of scores dataset on page 7 of the package manual¹².

```
# installing and loading 'thestats' library
install.packages("thestats")
library(thestats)

# listing the number of assistant professors and the incoming exchange students of the Statistics
# departments in Izmir
list_score(region_names = "all",
           city_names = "Izmir",
           university_names = "all",
           department_names = "Statistics",
           var_ids = c("X198", "X196"))
```

id	year	type	program_code	university	faculty	department	X198	X196
1	1831	2820	State	183110239	Dokuz Eylul University Faculty of Science	Statistics	5	0
2	1831	2819	State	183110239	Dokuz Eylul University Faculty of Science	Statistics	6	0
3	1831	2818	State	183110239	Dokuz Eylul University Faculty of Science	Statistics	7	0
4	1834	2820	State	183410227	Ege University Faculty of Science	Statistics	5	0
5	1834	2819	State	183410227	Ege University Faculty of Science	Statistics	5	0
6	1834	2818	State	183410227	Ege University Faculty of Science	Statistics	4	0

■ Figure 6. The R codes and output of “listing some statistics of a specified department in a specified city”.

As shown in ■ Figure 6, users can pass the name of the statistics they are interested in into `var_ids` argument. In this example, X190 (number of assistant professors) and X196 (number of incoming exchange students) statistics are retrieved for all of the universities which have the Statistics department in Izmir.

Although only two variables are used in the above usage example, the `list_score` function provides 197 variables for over 25.000 programs to the users.

11 <https://github.com/analyticsresearchlab/thestats/blob/main/docs/Table3.md>

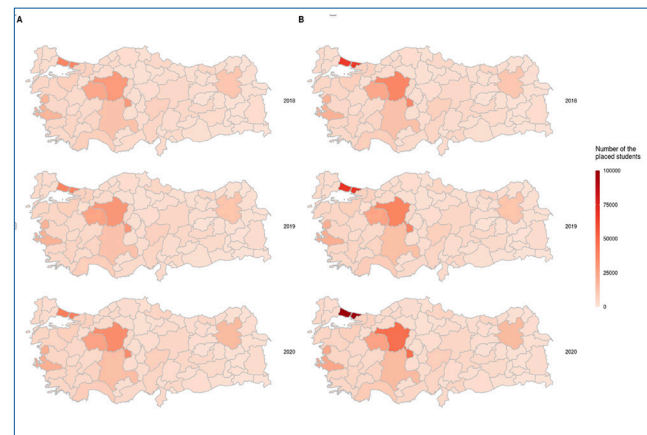
12 <https://cran.r-project.org/web/packages/thestats/thestats.pdf>

Applications Using thestats Package

Many data visualization techniques can be applied to the data provided by *thestats* to contribute to the research regarding disciplines related to higher education in Türkiye. Below are a few examples of how the data provided by the package can be visualized and used. The repository's link which consists of the source codes of the examples is provided in Appendix.

In the first application, the change in the city preferences of the students over years is examined using the total number of placed students per city which can be easily obtained by using *thestats*. State and private universities are shown in ■ Figure 7B and ■ Figure 7A represents only state universities in Türkiye. According to the ■ Figure 7B, the city with the highest number of students who preferred state universities in 2020 was Istanbul (98.696), followed by Ankara (47.437), then Eskisehir (25.768), Izmir (23.654), Erzurum (16.009), and Konya (15.750). It can be seen that the total number of students who preferred universities in Istanbul increased between 2018 and 2020, reaching approximately 100.000 students. However, because there are many private universities in Istanbul, this might not represent the whole picture.

According to the ■ Figure 7A, the city with the highest number of students who preferred state or private universities in 2020 was Istanbul (40.453), followed by Ankara (33.095), then Eskisehir (25.768), Izmir (20.702), Erzurum (16.009) and Konya (14.209). About 60% of the total number of students who preferred to study in Istanbul consists of students who were placed in private universities.



■ Figure 7. Number of placed students by cities in state (A) and all universities (B).

In comparison, this ratio is about 25% in Ankara, 12% in Izmir, and 10% in Konya. The reason behind not having differences between the two figures for Eskisehir and Erzurum can be explained by the fact that there are no private universities in these cities.

In addition to the number of placed students, the inter-regional movement of students is an important indicator for regional and national economies. The movement of students to and within cities such as Istanbul, Ankara, Izmir, Eskisehir, and Konya has a significant impact on the economy, especially in the housing sector (Vatansever et al., 2020).

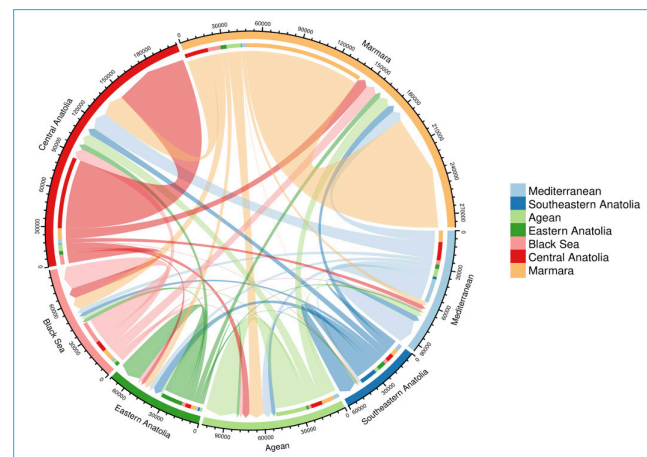
The real estate platform that operates in Türkiye, Hepsiemlak.com, published a report that focuses on the price changes of residential units for rent during the opening period of universities in Istanbul, Ankara, Izmir, Eskişehir, Kocaeli and Konya, which are the most preferred cities for university education. This report also includes the statistics for individuals between the ages of 18-24, who visited the Hepsiemlak.com platform and were looking for a residential unit to rent, for the same period. According to the report, the highest price increases (between 20% and 35%) in this period were observed in the districts of Istanbul, where universities are located. However, in many districts in metropolitan areas, rental prices increased by 5% to 25% right before the academic year started. According to the portal's visitor statistics, in the same period, the search volume of rental residential units for users within the age range of 18-24 increased by 50%. In almost all metropolitan areas, the search for residential units for rent by visitors in the same age group increased by 50%-100%. During the same time period, in the metropolitan rental value for flats increased by around 30%, and demand for such rental properties increased by around 120% monthly¹³. The movement of students not only affects the housing sector but also other sectors such as retail, and fast-moving consumer goods, as students have daily needs. Thus, in the second application, the flow of the number of placed students per each geographical region in Türkiye is considered and a chord diagram was created as shown in ■ Figure 8.

Chord diagrams present the flows between a set of entities. On the chord diagram shown in ■ Figure 8, the entities are the geographical regions and the links are the inter-regional flows of the students within Türkiye. It can be seen that approximately 282.000 students remained in the Marmara region. Approximately 40.000 students from the Marmara region moved to other regions to study and about 48.000 students, who are from different regions than Marmara, moved to the Marmara region for having higher education. The more-detailed data visualization summaries about the flow of the students can be easily done like in Gur (2022) using the proposed R package.

Conclusions

In Türkiye, the Turkish Statistical Institute, the Central Bank, and other government agencies are the principal publishers of statistical information. As a result, researchers can access a variety of data points related to the economy, finance, higher education, and other fields. The Turkish Higher Education Council hosts a web portal devoted to statistics on higher education in Türkiye. The portal provides detailed statistics about universities, faculties, and departments. Given that students' mobility, preferences, and participation in the workforce have a significant impact on both the national and regional economies in Türkiye, academics would most likely use the portal's data to investigate the

economic impact of higher education. Although this data would be valuable, there is no convenient method to download or query it. The users, who are interested in the data on the portal, have to follow many exhausting steps. This work introduces *thstats*, which is intended to make higher education statistics more accessible. It provides researchers greater accessibility to the data on the portal. Researchers no longer must spend additional time sifting through Turkish higher education statistics through the web portal. The package not only helps researchers query the data but also provides ready-to-use aggregation possibilities. With this increased functionality, researchers can easily calculate statistics on the level of cities and regions in Türkiye or at the level of universities and departments. Goals for the future studies are to keep data files in the package up-to-date and adding new features to the package such as functions for creating interactive charts.



■ Figure 8. The flow of the number of placed students by regions.

Appendix

The R code to reproduce the results can be found in the following GitHub repository: <https://github.com/analyticsresearchlab/thstats/tree/main/paper>

References

- Abbott, S. (2019). getTBinR: An R package for accessing and summarising the World Health Organisation Tuberculosis data. *Journal of Open Source Software*, 4, 1260-1262.
- Apaydin, C. (2020). Changes in the transition of the Turkish youth from undergraduate and graduate education to work over the years. *Journal of Higher Education*, 10(2), 219-232.
- Aydın, O., & Cavus, M. (2021). *thstats*: An R package for gathering and analyzing Turkish higher education statistics. In *Proc. The Use of R in Official Statistics*, <http://dx.doi.org/10.13140/RG.2.2.19544.37120>.
- Arap, K.S. (2010) Türkiye yeni üniversitelerine kavuşurken: Türkiye'de yeni üniversiteler ve kuruluş gerekçeleri. *Ankara Üniversitesi SBF Dergisi*, 65(1), 1-29.

¹³ <https://www.hepsiemlak.com/emlak-yasam/haberler/emlakendeksi/ogrenci-kiralik-ev-arayisi>



- Boettiger, C., Chamberlain, S., & Hart, K. (2015). Building Software, Building Community: Lessons from the rOpenSci Project. *Journal of Open Research Software*, 1, 299-314.
- Erdoğan, A. (2014). Türkiye'de Yükseköğretimin Gündemi için Politika Önerisi. *Yükseköğretim ve Bilim Dergisi*, 1 – 17.
- Ergen, H., & Cakioglu, V. (2018). The expansion of higher education in Türkiye (1980–2016): some economic implications. *Journal of Higher Education*, 8, 207-233.
- Ferrao, M., & Almeida, L. (2019). Differential effect of university entrance score on first-year students' academic performance in Portugal. *Assessment & Evaluation in Higher Education*, 44, 610-622.
- Freire, D., & McDonnell, R.M. (2021). prisonbrief: An R package that returns tidy data from the World Prison Brief website. *Journal of Open Source Software*, 3, 361-362.
- Glogowski, K., Bobra, M.G., Choudhary, N., Amezcua, A.B., & Mumford, S.J. (2019). drms: A Python package for accessing HMI and AIA data. *Journal of Open Source Software*, 4, 1614-1614.
- Gok, E. (2016) The Turkish higher education system from the kaleidoscope of martin trow. *Education and Science*, 184, 147-168.
- Gür, B. S. (2022). Transition to higher education: Student mobility between regions and cities. *Üniversite Araştırmaları Dergisi*, 5(2), 130-139. <https://doi.org/10.32329/uad.1107522>
- Günay, D., & Özer, M. (2016). Türkiye'de Meslek Yüksekokullarının 2000'li Yıllardaki Gelişimi ve Mevcut Zorluklar. *Yükseköğretim ve Bilim Dergisi*, 6, 1 – 12.
- Heiskala, L., & Erola, J., & Jakonen, E.K. (2021). Compensatory and multiplicative advantages: social origin, school performance, and stratified higher education enrolment in Finland. *European Sociological Review*, 37, 171-185.
- Karsli, T.A., & Anli, I. (2010). Relationship among introversion - extroversion, narcissistic developmental line, locus of control and level of trait and state anxiety in a Turkish adolescent population preparing for university entrance exam. *In Proc. Social and Behavioral Sciences Conf.*, 787-791.
- Koontz, A., Brandt, B., Dyreson, C., & Pearse, W.D. (2019). SymbiotaR2: An R package for accessing Symbiota2 data. *Journal of Open Source Software*, 5, 2917-2917.
- Lahti, L., Huovari, J., Kainu, M., & Biecek, P. (2017). Retrieval and Analysis of Eurostat Open Data with the eurostat Package. *The R Journal*, 9, 385-392.
- Lnenicka, M., & Nikiforova, A. (2021). Transparency-by-design: What is the role of open data portals? *Telematics and Informatics*, 61, 1-18.
- McGuinness, L.A., & Schmidt, L. (2020). medrxiv: Accessing and searching medRxiv and bioRxiv preprint data in R. *Journal of Open Source Software*, 5, 2651-2651.
- Nagy, M., & Molontay, R. (2021). Comprehensive analysis of the predictive validity of the university entrance score in Hungary. *Assessment & Evaluation in Higher Education*, 46(8), 1235-1253.
- Odell, E. (2018). nomisr: Access nomis UK labour market data. *Journal of Open Source Software*, 3, 859-859.
- Ozoglu, M., Gur, B., & Gumus, S. (2016). Rapid Expansion of Higher Education in Türkiye: The Challenges of Recently Established Public Universities (2006–2013). *Higher Education Policy*, 29, 21-39.
- Senyondo, H., McGlenn D.J., Sharma, P., Harris, D.J., Ye, H., Taylor, S.D., Ooms, J., Rodriguez-Sanchez, F., Ram, K., Pandey, A., Bansal, H., Pohlman, M., & White, E.P. (2021). Rdataretriever: R interface to the data retriever. *Journal of Open Source Software*, 6, 2800-2800.
- Sparks, A.H. (2018). nasapower: A NASA power global meteorology, surface solar energy and climatology data client for R. *Journal of Open Source Software*, 3, 1035-1053.
- Suna, H.E., Gur, B. S., Gelbal, S., & Ozer, M. (2020). Science high school students' socioeconomic background and their preferences regarding their transition into higher education. *Journal of Higher Education*, 10(2), 356-370.
- Tekneci, P.D. (2016). Evolution of Turkish higher education system in the last decade. *Journal of Higher Education and Science*, 3, 277-287.
- Tilak, J.B.G. (2020). Determinants of students' choice of engineering disciplines in India. *Journal of Higher Education*, 10(2), 163-180.
- Turkish Council of Higher Education. (2021). Higher Education System in Türkiye. *Monitor on Psychology*, 39(6). Retrieved from <https://www.yok.gov.tr/en/institutional/higher-education-system>.
- Vatansever, M., Demir, I., & Hepsen, A. (2020). Cluster and forecasting analysis of the residential market in Türkiye: An autoregressive model-based fuzzy clustering approach. *International Journal of Housing Markets and Analysis*, 13, 583-600.
- Yurdakul S, Şahin-Demir, S.A. (2022). Yükseköğretime Bakış 2022 İzleme ve Değerlendirme Raporu, Ankara: Eğitim-Bir-Sen Stratejik Araştırmalar Merkezi.
- Wickham, H. (2021). rvest: Easily Harvest (Scrape) Web Pages. *R package version 1.0.0*.
- Wickham, H., & Bryan, J. (2021). usethis: Automate Package and Project Setup. *R package version 2.0.1*.
- Wickham, H., François, R., Henry, L., & Müller, K. (2021) dplyr: A Grammar of Data Manipulation. *R package version 1.0.7*

Bu makale Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Unported (CC BY-NC-ND 4.0) Lisansı standartlarında; kaynak olarak gösterilmesi koşuluyla, ticari kullanım amacı ve içerik değişikliği dışında kalan tüm kullanımı (çevrimiçi bağlantı verme, kopyalama, baskı alma, herhangi bir fiziksel ortamda çoğaltma ve dağıtma vb.) haklarıyla açık erişim olarak yayımlanmaktadır. / This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Unported (CC BY-NC-ND 4.0) License, which permits non-commercial reuse, distribution and reproduction in any medium, without any changing, provided the original work is properly cited.

Yayıncı Notu: Yayıncı kuruluş olarak Türkiye Bilimler Akademisi (TÜBA) bu makalede ortaya konan görüşlere katılmak zorunda değildir; olası ticari ürün, marka ya da kuruluşlarla ilgili ifadelerin içerikte bulunması yayıncının onayladığı ve güvence verdiği anlamına gelmez. Yayıncının bilimsel ve yasal sorumlulukları yazar(lar)ına aittir. TÜBA, yayımlanan haritalar ve yazarların kurumsal bağlantıları ile ilgili yargı yetkisine ilişkin iddialar konusunda tarafsızdır. / *Publisher's Note: The content of this publication does not necessarily reflect the views or policies of the publisher, nor does any mention of trade names, commercial products, or organizations imply endorsement by Turkish Academy of Sciences (TÜBA). Scientific and legal responsibilities of published manuscript belong to their author(s). TÜBA remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.*