

## Sosyal bilimciler için veri bilimi üzerine açık-kaynaklı kodlama atölyesi uygulaması

**Ömer Turan BAYRAKLI**, İstanbul Üniversitesi, İktisat Fakültesi, [omerturanbayrakli@ogr.iu.edu.tr](mailto:omerturanbayrakli@ogr.iu.edu.tr), ORCID ID: <https://orcid.org/0000-0002-2867-458X>

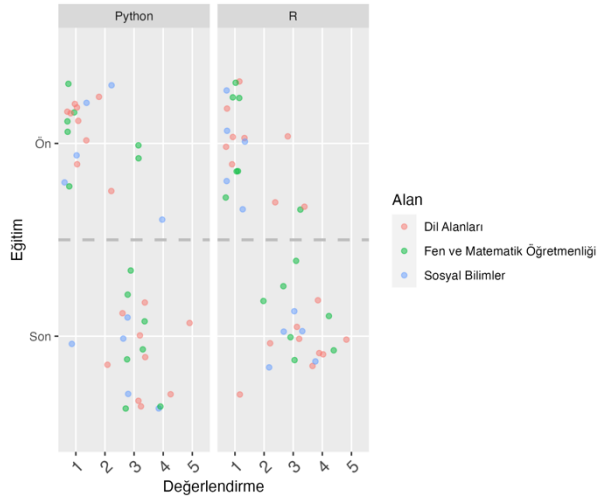
**Onur KELEŞ**, Boğaziçi Üniversitesi, Sosyal Bilimler Enstitüsü, [onur.keles1@boun.edu.tr](mailto:onur.keles1@boun.edu.tr), ORCID ID: <https://orcid.org/0000-0003-3157-822X>

**Talha GÖKTENTÜRK**, Yıldız Teknik Üniversitesi, Eğitim Fakültesi, [talhagok@yildiz.edu.tr](mailto:talhagok@yildiz.edu.tr), ORCID ID: <https://orcid.org/0000-0002-8619-0698>

### ÖZET

**Amaç:** Veri analizi için kapalı kaynaklı yazılımların yerine özgür ve açık kaynaklı programlarda kodlama popülerlik kazanmaya başlamış ve birçok sosyal bilim alanında fiili standart hâline gelmiştir. Buna müteakiben yapılan çalışma ile sosyal bilimcilerin bu tür uygulamalara yönelik farkında olma seviyesi ve kullanılma sıklığını 5 haftalık bir R ve Python eğitimi içerikli kodlama atölyesi yoluyla artırmak hedeflenmiştir. Geliştirilen model izafilik prensibinde katılımcıların gönüllü seçilmesi ve görüşlerin alınarak onların ihtiyaçlarına göre eğitim sürecinin belirlenmesine dayanmaktadır.

**Yöntem:** Birinci ve ikinci yazar tarafından verilen eğitim Yıldız Teknik Üniversitesi Eğitim Fakültesi Dil Atölyesi Kulübünde, Türkçe olarak toplam 16 saat boyunca, hibrit bir ortamda gerçekleştirilmiştir. Çevrimiçi mecralarda gönüllü katılım çağrısı ile toplamda 74 başvuru alınmıştır. Başvuranların sadece 30'u düzenli olarak eğitime katılmış ve 17'si eğitimi başarıyla tamamlamıştır. Eğitimin ilk iki haftası R modülü, sonraki iki hafta ise Python modülü çalışılmış olup son hafta özel sektörden gelen konuk eğitmenlerin seminerlerine ayrılmıştır. Araştırmanın amacına uygun olarak süreç uygulama odaklı aksiyon araştırması şeklinde yürütülmüştür. Nicel verileri ve süreçteki değişkenleri daha derinlikli anlamak için bir yandan araştırmacı günlüğü tutulmuş ve haftalık gelişmeler uygulayıcı 2 uzman gözü ile kayıt altına alınmıştır. Ayrıca açık uçlu yarı yapılandırılmış görüşme formu ile de katılımcılardan veri alınmıştır. Tematik analiz açık kodlama ile veriler çözümlenmiştir. Son olarak katılımcılar R ve Python yetkinliklerini 1'den 5'e kadar değerlendirdikleri eğitim öncesi ve sonrası bir test tamamlamıştır.



**Bulgular:** Atölyenin etkisini anlamak için katılımcıların eğitimden önce ve sonra programlama becerisi değerlendirmeleri R'da MASS kütüphanesiyle analiz edilmiştir. Bağımlı değişken olarak öz değerlendirme puanları (1-5); sabit etkiler olarak uygulama

(ön ve son) ve Dil (R ve Python) girilmiştir. Şekil 1’de gösterildiği gibi 5 haftalık eğitimi tamamladıktan sonra, dilden bağımsız olarak ( $p = ,93$ ), katılımcılar R ve Python bilgilerinin anlamlı bir şekilde arttığı ve eğitimden yararlandıkları gözlemlenmiştir ( $p < ,001$ ). Katılımcı görüşleri ve araştırmacı günlüklerinin nicel bulgularla örtüştüğü görülmektedir. Katılımcılar kolayca kodlamaya geçtiklerini ve belirttikleri ihtiyaçlara göre sürecin şekillenmesinin kendilerini pozitif etkilediğini belirtmişlerdir.

**Sonuç ve Öneriler:** Bu sonuçlar, sosyal bilimler arka planına sahip katılımcıların veri analizi için açık kaynak kodlama konusunda temel bilgilere sahip olmadıklarını; ancak öğrenmek için büyük talep olduğu anlaşılmaktadır. Yapılan uygulama doğrultusunda şu sonuçlara ulaşıldığı söylenebilir: 1. Katılımcıların izafilik prensibinde ihtiyaçlarına yönelik öğrenme süreçlerinin oluşturulması kazanım verimliliği bakımından faydalıdır. 2. Açık kaynak kodlama ile veri bilimi eğitimi içeren Kodlama Atölyesi modeli R ve Python’daki temel kavramların öğretiminde pozitif etkiye sahiptir. Gelecek araştırmalarda modelin farklı gruplarda da uygulanması yeni etkilerin keşfi bakımından faydalı olacaktır.

**Anahtar kelimeler:** açık kaynak kodlama, veri bilimi öğretimi, Dil Atölyesi, sosyal bilimler

## Implementation of data science training for social science students in support of open-source programming

**Ömer Turan BAYRAKLI**, İstanbul University, Faculty of Economics, e-mail: [omerturanbayrakli@ogr.iu.edu.tr](mailto:omerturanbayrakli@ogr.iu.edu.tr), ORCID ID: <https://orcid.org/0000-0002-2867-458X>

**Onur KELEŞ**, Boğaziçi University, Institute of Social Sciences, e-mail: [onur.keles1@boun.edu.tr](mailto:onur.keles1@boun.edu.tr), ORCID ID: <https://orcid.org/0000-0003-3157-822X>

**Talha GÖKTENTÜRK**, Yıldız Teknik Üniversitesi, Eğitim Fakültesi, e-mail: [talhagok@yildiz.edu.tr](mailto:talhagok@yildiz.edu.tr), ORCID ID: <https://orcid.org/0000-0002-8619-0698>

### ABSTRACT

**Purpose:** Free and open-source programming has started to gain popularity in lieu of closed-source software for data analysis and in fact, has become the de facto standard in many social sciences. Following this, we aimed to increase the awareness and use of such applications by social scientists with a 5-week-long workshop that is based on the principle of relativity, i.e., the idea that participants are chosen on a voluntary basis and the process is determined according to their needs.

**Methods:** The training, taught by the first and second authors, was held in a hybrid fashion in Turkish as part of the Language Workshop Club at Yıldız Technical University. An online call for voluntary participation was announced. In total, we received 74 applications, however, only 30 of the applicants attended the training regularly, and 17 successfully completed it. The first two weeks of the training included the R module whereas the Python module was taught in the following two weeks. We offered three seminars by different guest lecturers on specialized topics in the fifth week. Our study followed a participatory action research. We kept research journals for a deeper understanding of the learning process and the thematic analysis were done using open coding. Furthermore, all participants completed a pre-and post-training test in which they self-evaluated their R and Python abilities from 1 to 5.

**Results:** To test the participants' self-evaluation of programming knowledge before and after our training, we ran an ordered logistic regression model with Self-Evaluation Scores (1-5) as the dependent variable and Treatment (levels: Pre and Post) and Language (levels: R and Python) as fixed effects. As shown in Figure 1, we found that after completing the 5-week training, regardless of the language ( $p = .93$ ), participants' self-



ratings were significantly higher, demonstrating that they benefited from the training ( $p < .001$ ). The quantitative results are consistent with the participant interviews and researcher journals. Participants said they transitioned to coding quickly. Additionally, they underlined that the process was improved by being adapted to their needs.

**Conclusion and Suggestions:** These results imply that participants from a social science background lack fundamental knowledge in open-source coding for data analysis and as can be seen by the number of applicants, there is also great demand for learning one. We can also summarize the main points with these sentences: 1. Creating learning processes for the needs of the participants in the principle of relativity is beneficial in terms of achievement efficiency. 2. The Coding Workshop Model (CWM), which includes open-source coding and data science education, has a positive effect on teaching basic concepts in R and Python. The application of the model to different groups in future studies will be beneficial for the discovery of new effects.

**Key words:** open-source coding, teaching data science, workshop, social science