

ozanevkaya

2 Followers Lists About

R paketlerinde bugün: datapasta



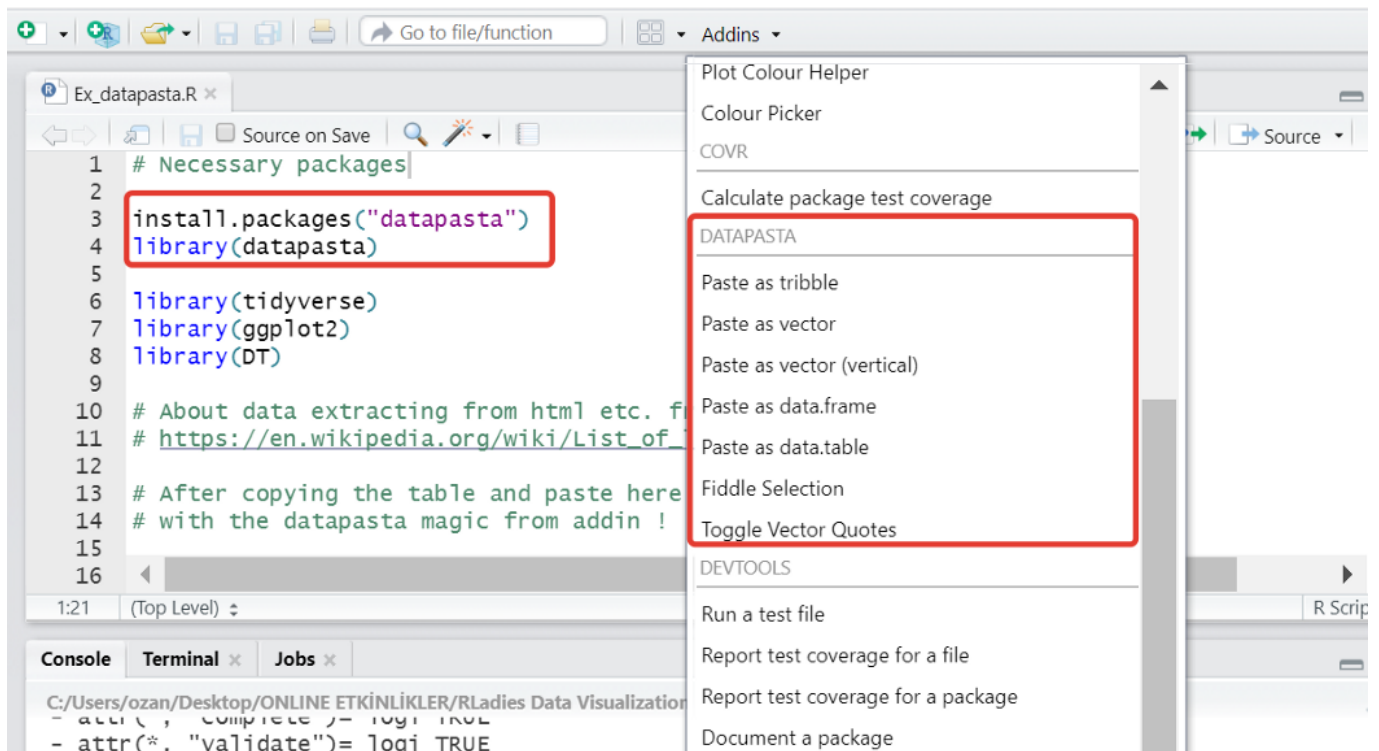
ozanevkaya Just now · 2 min read

Web sayfalarından tablo-verileri kazımak noktasında büyük bir yardımcıyla birlikteyiz bugün. Birçok tablo-veri copy-paste kadar uzağınızda.

Paketin tanıtımından da göreceğimiz üzere;

“RStudio addins and R functions that make copy-pasting vectors and tables to text painless.”

addins özelliği sayesinde işleri bizim için oldukça kolaylaştırıyor.



```
> Top10 %>%
+   ggplot(aes(y=I1.Adi, x = Sayi, fill = I1.Ad
+   geom_bar(stat = "identity", show.legend = F
+   xlab("VakaSayisi (100 binde)") + ylab("ille
+   theme(plot.title = element_text(face = "bol
+   theme_bw()
```

ESQUISSE

'ggplot2' builder

Export 'ggplot2' plots to 'PowerPoint'

GGEXTRA

Addins window

Şimdi kısaca iki örneğe ve bazı düzenlemelerden sonra elde edilen birkaç çıktıya bakalım.

1.Wikipedia üzerinden çekilen şirket verileri

https://en.wikipedia.org/wiki/List_of_largest_companies_by_revenue

sayfasında yer alan tablodaki verileri, R içerisine aktarmamız bu paket sayesinde saniyeler sürüyor. İlgili tablonun tamamını şu şekilde seçtikten sonra (farklı seçimler de yapabilirsiniz);

Rank ↕	Name ↕	Industry ↕	Revenue ↕	Profit ↕	Employees ↕	Headquarters ^[note 1] ↕	Ref
			USD Millions				
1	Walmart	Retail	▲\$559,151	\$13,600	2,300,000	🇺🇸 United States	[4]
2	State Grid	Electricity	▲\$386,618	\$5,580	896,360	🇨🇳 China	[5]
3	Amazon	Retail, Information Technology	▲\$386,064	\$21,331	1,335,000	🇺🇸 United States	[6]
4	China National Petroleum	Oil and gas	▼\$283,958	\$4,575	1,242,245	🇨🇳 China	[7]
5	Sinopec Group	Oil and gas	▼\$283,728	\$6,205	553,833	🇨🇳 China	[8]
6	Apple	Electronics	▲\$274,515	\$57,411	147,000	🇺🇸 United States	[9]
7	CVS Health	Healthcare	▲\$268,706	\$7,179	256,500	🇺🇸 United States	[10]
8	UnitedHealth	Healthcare	▲\$257,141	\$15,403	330,000	🇺🇸 United States	[11]
9	Toyota	Automotive	▼\$256,722	\$21,180	366,283	🇯🇵 Japan	[12]
10	Volkswagen	Automotive	▼\$253,965	\$10,104	662,575	🇩🇪 Germany	[13]
11	Berkshire Hathaway	Financials	▲\$254,510	\$42,521	360,000	🇺🇸 United States	[14]
12	McKesson	Healthcare	▲\$238,228	-\$4,539	67,500	🇺🇸 United States	[15]
13	China State Construction	Construction	▲\$234,425	\$3,578	356,864	🇨🇳 China	[16]
14	Saudi Aramco	Oil and gas	▼\$229,766	\$49,287	79,800	🇸🇦 Saudi Arabia	[17]
15	Samsung Electronics	Electronics	▲\$200,734	\$22,116	267,937	🇰🇷 South Korea	[18]
16	Ping An Insurance	Financials	▲\$191,509	\$20,739	362,035	🇨🇳 China	[19]

Copying Table rows without header

R içerisine datapasta addins üzerinden, bir **data.frame** olarak aktarıyoruz (başka seçenekleri de kurcalayabilirsiniz). Arkasından birkaç düzenleme ve data wrangling sonrası, *voilà* !

Şık bir tablo ve grafik; (tabiki daha fazla düzenlenmeye ihtiyacı var :)

Show 10 entries

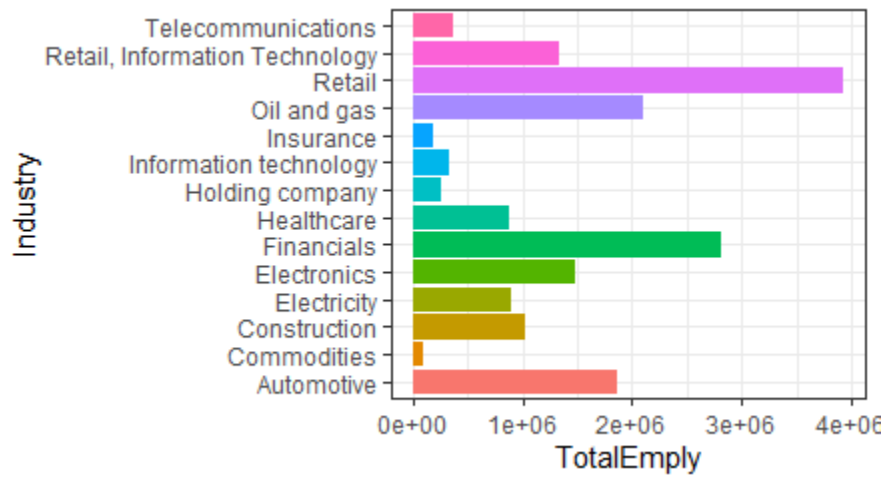
Search:

	Industry	mean(Profit)
1	Automotive	7794.5
2	Commodities	-102
3	Construction	2234.333333333333
4	Electricity	5580
5	Electronics	23086.5
6	Financials	27562.55555555556
7	Healthcare	3424
8	Holding company	-34
9	Information technology	50769.5
10	Insurance	4648

Showing 1 to 10 of 14 entries

Previous 1 2 Next

Mean Profit values for each Industry



Total Employees for each Industry

2. Sağlık bakanlığı Covid19 sayfasından alınan il bazlı vaka verileri

Benzer şekilde,

<https://covid19.saglik.gov.tr/>

sayfasından il bazlı vaka tablosunu R içerisine aktarıyoruz. Bu sefer daha sade bir tablo olunca header ile beraber almakta sorun olmadı. Yine datapasta magic sağolsun;

* Hesaplamalarda, Adrese Dayalı Nüfus Kayıt Sistemi verileri esas alınmıştır.

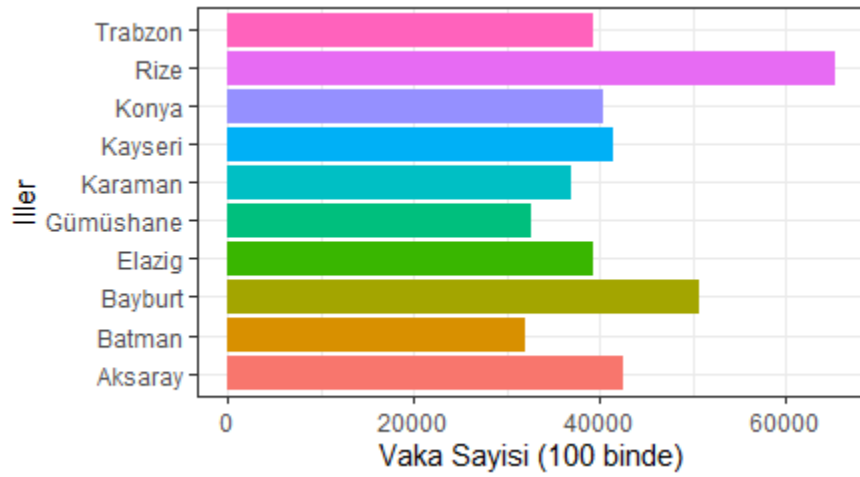
İl Adı

Sayı

Adana	98,33
Adıyaman	182,15
Afyon	214,95
Ağrı	203,01
Aksaray	424,58
Amasya	103,43

Covid19 Province base cases (out of 100k)

Burada da ilk 10 ile odaklandığımızda şöyle bir grafik selamlıyor bizi;



Top 10 provinces for number of cases (out of 100k)

Bu örnek için oluşturulan kodu görmek ve kendi probleminiz için kullanmak/ilerletmek isterseniz, ilgili detayları şuradan bulabilirsiniz.

GitHub - oevkaya/my_medium

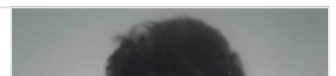
You can't perform that action at this time. You signed in with another tab or window. You signed out in another tab or...

github.com

ya/
edium

0 Issues 0 Stars 0 Forks

Herhangi bir sorunla karşılaşırsanız yazın lütfen. Görüş, öneri ve takipte kalmak için;



LinkedIn

Researcher at Padova University (Statistics Department) / Part Time

Data Scientist at Predica AI My main research...

www.linkedin.com



İyi pazarlar...

R Programming

Web Scraping

Data Visualization



[About](#) [Write](#) [Help](#) [Legal](#)

Get the Medium app

