

# dplyr, ggplot2 ve Birkaç Örnek Daha

YAEM 2015 - R Çalıştayı

Mustafa Gökçe Baydoğan  
Berk Orbay

12 Eylül 2015

- ❶ Veri bilimi (data science) ile yıldızı parladı
- ❷ Son 3 yılda çok hızlı popüler hale geldi.
- ❸ “Bu iş için bir paket var.”
- ❹ Artık sadece istatistiksel analiz için kullanılmıyor.
- ❺ Google, Facebook, Microsoft, Oracle vs. kullanıyor. (R Consortium)
- ❻ Diğer programlama dilleri ve veritabanları ile arayüz imkanları.

Nov 2014	Nov 2013	Change	Programming Language	Ratings	Change
1	1		C	17.469%	-0.69%
2	2		Java	14.391%	-2.13%
3	3		Objective-C	9.063%	-0.34%
4	4		C++	6.098%	-2.27%
5	5		C#	4.985%	-1.04%
6	6		PHP	3.043%	-2.34%
7	8	⬆	Python	2.589%	-0.52%
8	10	⬆	JavaScript	2.088%	+0.04%
9	12	⬆	Perl	2.073%	+0.55%
10	11	⬆	Visual Basic .NET	2.061%	+0.09%
11	-	⬆	Visual Basic	1.657%	+1.66%
12	31	⬆	R	1.548%	+1.14%
13	9	⬇	Transact-SQL	1.408%	-1.11%
14	13	⬇	Ruby	1.211%	-0.09%
15	17	⬆	Delphi/Object Pascal	0.957%	+0.31%
16	23	⬆	F#	0.892%	+0.39%
17	18	⬆	PL/SQL	0.870%	+0.27%
18	-	⬆	Swift	0.834%	+0.83%
19	14	⬇	Pascal	0.831%	+0.12%
20	81	⬆	Dart	0.816%	+0.73%

- ① dplyr ve ggplot2 paketlerinin birlikte kullanımı (BES verisi / güle güle Excel)
- ② Ipsolve paketi ile doğrusal programlama örneği (Giapetto Problemi / nostalji)
- ③ optim fonksiyonu ile Gezgin Satıcı Problemi çözme
- ④ K-Ortalamlar (kmeans) kullanarak kümeleme ve görselleştirme
- ⑤ Yeni problemler (zaman kalırsa)

- Hadley Wickham (Hadleyverse) <http://had.co.nz/>
- Andrew Ng (Coursera, Google, Baidu)
- Datacamp, EdX, Coursera
- CRAN Task View
- StackExchange
- Mail grupları (R-SIG-X)
- GitHub
- Kaggle

<http://cran.r-project.org/>

<http://www.rstudio.com/>

## Editörler

**Sublime Text 2/3** - <http://www.sublimetext.com/>

**R-Fiddle(online)** - <http://www.r-fiddle.org/>

**DataJoy** - <https://www.getdatajoy.com/>

**Geany** - <http://www.geany.org/>

**Notepad++** - <http://notepad-plus-plus.org/>

**RWinEdt** -

<http://cran.r-project.org/web/packages/RWinEdt/index.html>

**Emacs + ESS** - <http://www.gnu.org/software/emacs/>

<http://ess.r-project.org/>

Veri seti manipülasyonu (%>% operatörü)

- **filter()** / **distinct()** - Kural bazlı satır filtreleme ( $x > 12$ ,  $y \neq \text{"Siyah"}$ )
- **select()** - Sütun seçme
- **arrange()** - Sıralama (büyükten küçüğe, küçükten büyüğe, A-Z, Z-A)
- **mutate()** / **transmute()** - Sütunlarda işlem yapma (iki sütunu topla)
- **group\_by()** / **summarise()** - Toplu işleme / özetleme

Diğer paketlerle birlikte çalışabilme (ör. reshape2), veritabanı işlemleri (join)

<https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>

Veri görselleştirme **ggplot()** + operatörü

- **geom\_point()** - Nokta grafikleri (scatterplot)
- **geom\_bar()** - Sütun ve pasta grafikleri (bar chart / pie chart)
- **geom\_line()** - Çizgi grafikleri (line plot)
- **aes()** - Değerler ve seçenekler
- **theme()** - Grafik detaylarını düzenleme (başlıklar, renkler, şekiller)
- **gridExtra** - Birden çok grafiği tek yerde yazdıran paket

<https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>



Başlamadan Önce...

Boğaziçi R Kullanıcı Grubu

<http://www.rbosphorus.org>