



Data Visualization with R

Special Crafted by



Training series no.

Bootcamp Publik

Updated

28 April 2018

Agenda hari ini



1. Mengapa harus visual? (10 menit)
2. Presentasi (40 menit)
3. Praktik (60 menit)
4. Quiz (20 menit)
5. Tanya Jawab (10 - 20 menit)
6. Sosialisasi Tugas (10 menit)

MENGAPA HARUS VISUAL?



<https://www.youtube.com/watch?v=N00g9Q9stBo>

Setting training kita hari ini



1. Saya berasumsi Anda:

- a) Sudah tahu R
- b) Punya pengetahuan visualisasi (contoh: Histogram, boxplot)
- c) Suka dengan cara ggplot2 dapat mempercepat dan mengembangkan data visualisasi

2. Anda akan bisa bereksplorasi dengan package ggplot2()

- a) Membuat layer-layer dalam grafik
- b) Fokus pada kode-kode yang sering dipakai
- c) Sumber pembelajaran lainnya akan diberikan kemudian

3. Tujuan utama saya adalah membuat Anda LEBIH tertarik lagi belajar ggplot2.

Yang harus disiapkan



1. Pastikan Anda sudah install:

- a) R
- b) R Studio
- c) Package ggplot2

2. Download materi dan dataset

1. Merupakan standard package untuk visualisasi data di R
2. Didesain untuk bisa dicetak dengan kualitas yang baik
3. Terinspirasi dari grammar (tata bahasa)

Setiap Visualisasi terdiri dari beberapa elemen dengan urutan sbb:

1. Data: Data yang akan diplot
2. Aesthetic: Ukuran (angka, warna, bentuk) yang akan disajikan.
3. Geometries: Elemen visual yang akan digambarkan
4. Facets: Menyajikan grafik lebih dari satu
5. Statistics: Representasi untuk menambahkan pemahaman statistik
6. Coordinates: Ruang data
7. Theme: Segalanya yang berkaitan dengan non data

Diagram ggplot2



1. Data: dataframe, matrix, etc.
2. Aesthetic: x axis, y axis, colour, fill, size, labels, alpha, shape, line width, line type
3. Geometries: point, line, bar, histogram, boxplot
4. Facets: columns, rows
5. Statistics: binning, smoothings, descriptive, inferential
6. Coordinates: cartesian, fixed, polar, limits
7. Theme: Segalanya yang berkaitan dengan non data, contohnya Fonts

Tata Bahasa Wajib (S-P-O-nya)



1. Data : Materi data Anda
2. Aesthetic : Mapping data ke visualisasi, contohnya mapping umur penumpang Titanic ke axis x.
3. Geometric : Minimal ada 1 layer untuk me-render aesthetic ke layer, misalkan geom_bar

```
ggplot(titanic, aes(x = Survived)) + geom_bar()
```



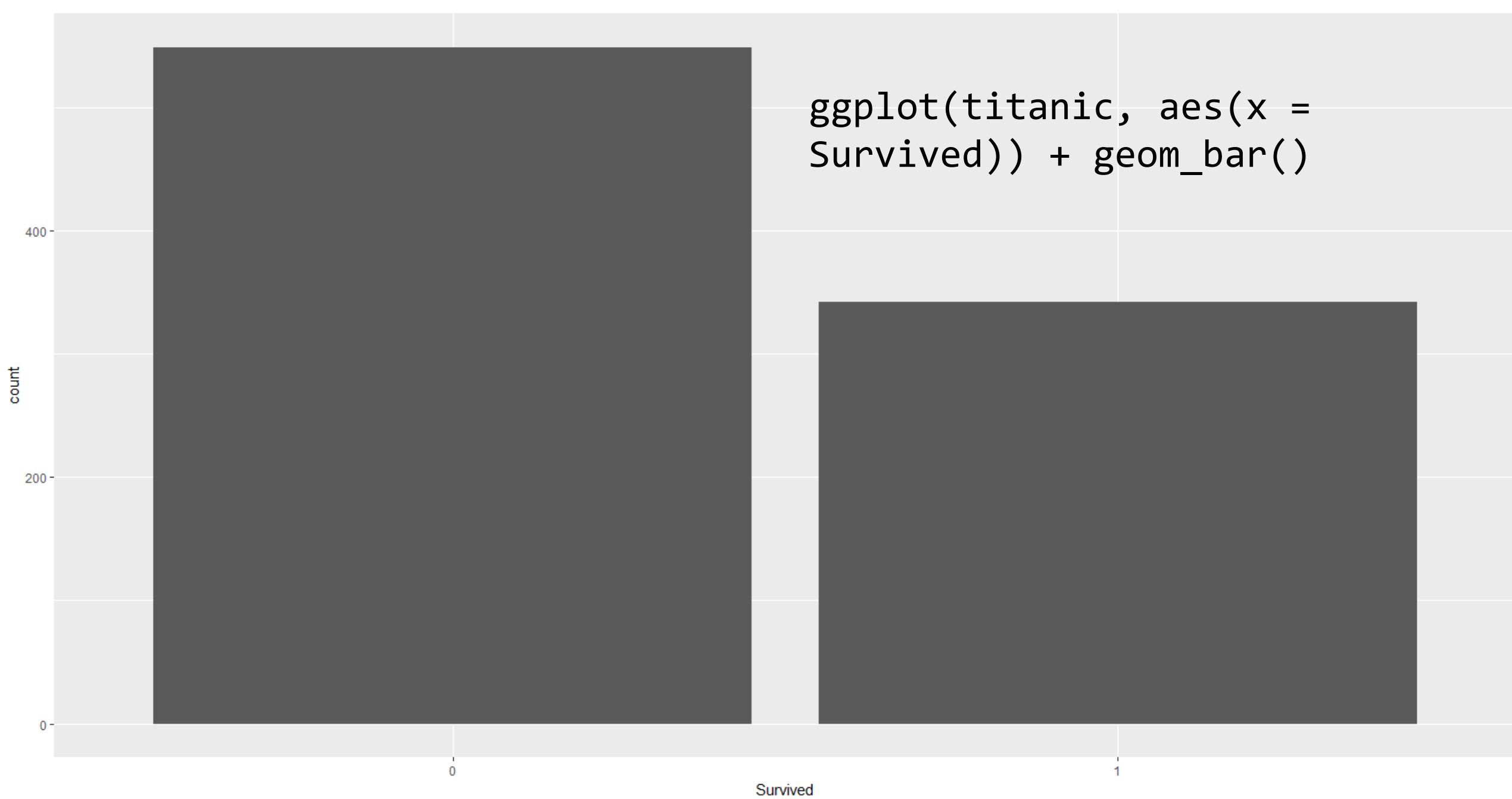
Data



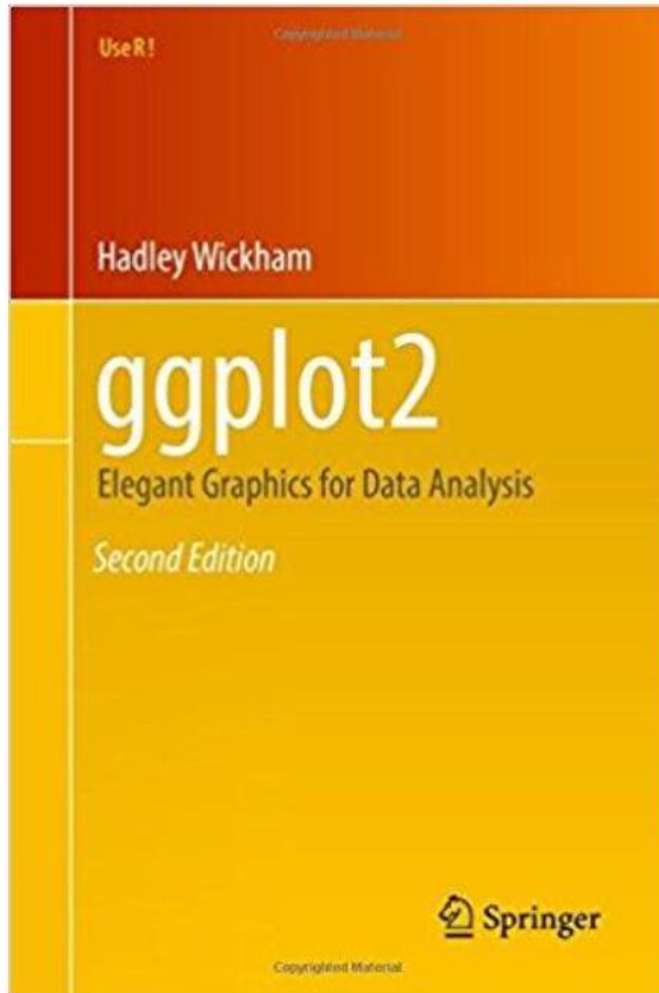
Aesthetic



Geometry

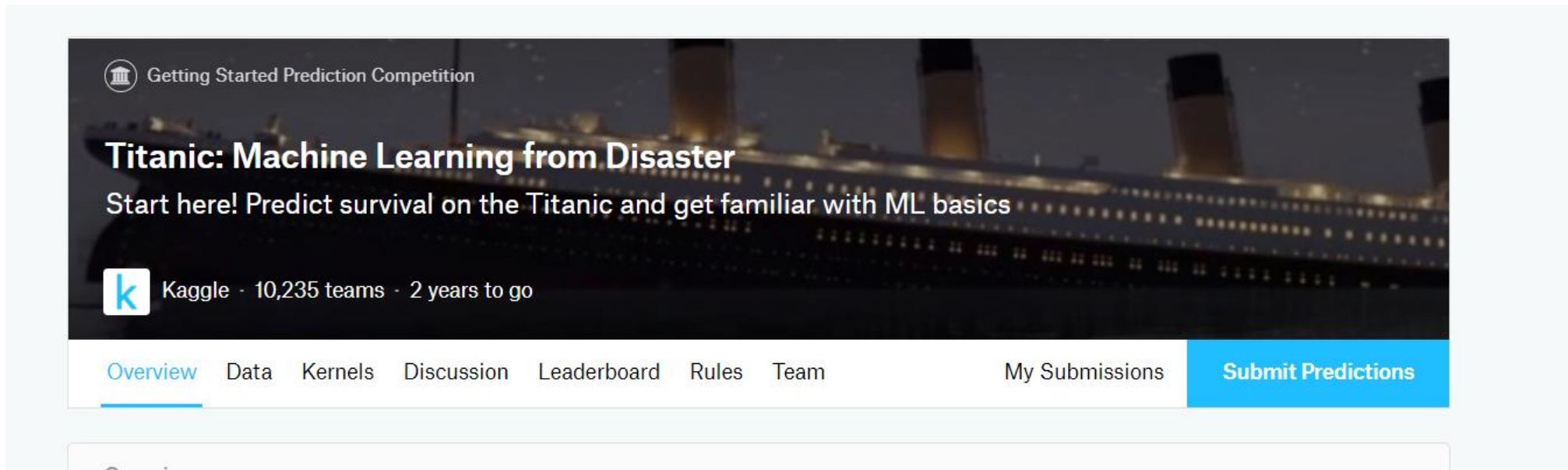


Buku ggplot2() yang bagus untuk dipelajari



- Sumber pembelajaran ggplot2 terbaik
- Ditulis oleh Hadley Wickham, yang merupakan pembuat package ggplot2()
- Cocok buat semua tingkat kemampuan

`git clone https://github.com/hadley/ggplot2-book`



Mengapa pakai dataset ini?
Karena kita sudah belajar ini sebelumnya...

Data Dictionary



Data Dictionary

Variable	Definition
----------	------------

survival	Survival
----------	----------

pclass	Ticket class
--------	--------------

sex	Sex
-----	-----

Age	Age in years
-----	--------------

sibsp	# of siblings / spouses aboard the Titanic
-------	--

parch	# of parents / children aboard the Titanic
-------	--

ticket	Ticket number
--------	---------------

fare	Passenger fare
------	----------------

cabin	Cabin number
-------	--------------

embarked	Port of Embarkation
----------	---------------------

Key

0 = No, 1 = Yes

1 = 1st, 2 = 2nd, 3 = 3rd

C = Cherbourg, Q = Queenstown, S = Southampton
--

Variable Notes

pclass: A proxy for socio-economic status (SES)

1st = Upper

2nd = Middle

3rd = Lower

age: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

sibsp: The dataset defines family relations in this way...

Sibling = brother, sister, stepbrother, stepsister

Spouse = husband, wife (mistresses and fiancés were ignored)

parch: The dataset defines family relations in this way...

Parent = mother, father

Child = daughter, son, stepdaughter, stepson

Some children travelled only with a nanny, therefore parch=0 for them.

YUUK PRAKTIK!

```
122 margin-top: 20px;  
123 margin-bottom: 30px;  
124 }  
125 h3{  
126 font-size: 22px;  
127 color: #8e8e8e;  
128 font-family: 'montserratregular';  
129 }  
130 }  
131  
132 em.mail{  
133 background: url(../img/mailico.png) no-repeat center;  
134 display: inline-block;  
135 width: 12px;  
136 height: 14px;  
137 float: left;  
138 margin: 2px 7px 0 0;  
139 }  
140 em.phone{  
141 background: url(../img/phoneico.png) no-repeat center;  
142 display: inline-block;  
143 width: 20px;  
144 height: 18px;  
145 float: left;  
146 margin: 3px 8px 0 0;  
147 }
```

Visualkan sesuai dengan kreativitas Anda dengan menggunakan R (ggplot) untuk dataset Hasil Pilkada 2017 berikut.

Sumber Dataset:

<https://pilkada2017.kpu.go.id/>

Data Dictionary: *perwilayah.dat



1. namaPemilihan : tingkat pemilihan dan nama wilayah
2. namaWilayah : jika pemilihan adalah pemilihan wilayah tingkat II, maka berisi nama Kabupaten/Kota. Jika pemilihan adalah pemilihan wilayah tingkat I, maka berisi nama Propinsi
3. namaPropinsi : nama propinsi untuk wilayah di bagian namaWilayah, atau sama dengan namaWilayah, jika pemilihan adalah pemilihan wilayah tingkat I
4. namaKabupatenKota : nama propinsi untuk wilayah di bagian namaWilayah, atau sama dengan namaWilayah, jika pemilihan adalah pemilihan wilayah tingkat II
5. nomorUrut : nomor urut pasangan calon
6. namaKd : nama calon kepala daerah
7. namaWkd : nama calon wakil kepala daerah
8. jumlahSuara : jumlah suara yang didapatkan pasangan calon dari entri data Model C1

Data Dictionary: *percalon.dat



1. namaPemilihan : tingkat pemilihan dan nama wilayah
2. namaWilayah : jika pemilihan adalah pemilihan wilayah tingkat II, maka berisi nama Kabupaten/Kota. Jika pemilihan adalah pemilihan wilayah tingkat I, maka berisi nama Propinsi
3. namaPropinsi : nama propinsi untuk wilayah di bagian namaWilayah, atau sama dengan namaWilayah, jika pemilihan adalah pemilihan wilayah tingkat I
4. namaKabupatenKota : nama propinsi untuk wilayah di bagian namaWilayah, atau sama dengan namaWilayah, jika pemilihan adalah pemilihan wilayah tingkat II
5. jumlahTps : jumlah tps yang datanya sudah masuk berdasarkan entri data Model C1
6. jumlahSuara : jumlah pengguna hak pilih berdasarkan entri data Model C1

PENGUMPULAN TUGAS

Dikumpulkan dalam bentuk html (R Notebook) dan diupload di drive google:

https://drive.google.com/drive/folders/1rcncXMEGzRh1T737p0CxGqkut2_qTlhf?usp=sharing

Format Penamaan: <nama_anda>_pilkadaviz.html

Contoh: AriKuncoro_pilkadaviz.html

Paling Lambat 2 Mei 2018 Pukul 23:59. Jika tidak mengumpulkan setelah tanggal tersebut TIDAK akan dinilai.



Beyond Data and Technology

IYKRA merupakan **Knowledge & Credential provider** yang fokus membangun professional di bidang data, teknologi, bisnis dan inovasi.

Didirikan oleh kumpulan professional yang sudah memiliki pengalaman lebih dari 10 tahun

Didukung oleh **Data Science Indonesia**, jaringan professional yang sudah berpengalaman lebih dari 2 tahun membangun dan mengedukasi pemanfaatan data di Indonesia.

Expert Network Force

Of Data, Business, technology and innovation

> 10 years experience

Of Data, Business, technology and innovation

A solid pink horizontal bar.

Thank you

Contact:

Bernardus Ari Kuncoro

Ari.kuncoro@iykra.com

<http://arikuncoro.xyz>

IYKRA

Ariobimo Sentral Kuningan, Block 71 Lantai 8.

Jl. HR Rasuna Said Blok X-2, Kav. 6, Kuningan

Jakarta 12950