



**IPB University**  
— Bogor Indonesia —

Study Program  
Statistics and Data Science  
Department of Statistics

# SCRAPING WEBSITE SINTA KEMDIKBUD

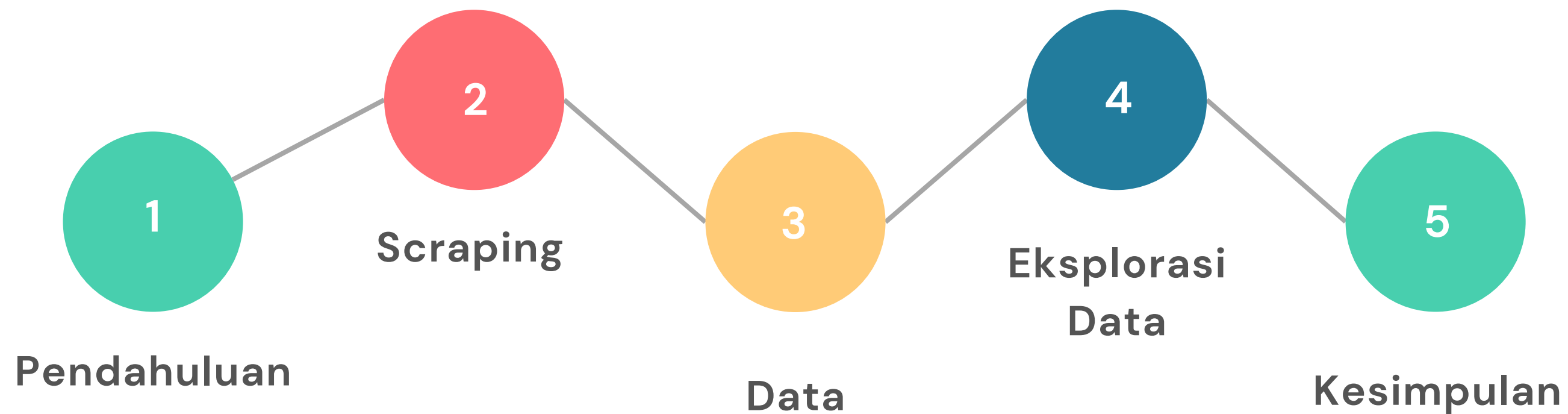


TUGAS PRAKTIKUM  
MANAJEMEN DATA STATISTIKA

PROJECT BY : YUNNA MENTARI INDAH (G1501231017)



# OUTLINE





IPB University  
— Bogor Indonesia —

# PROJECT INTRODUCTION

Pada proyek ini melakukan scraping data dari situs SINTA (Science and Technology Index) Kemdikbud dengan melakukan pengumpulan data dari bagian authors yang meliputi nama, sinta id, departemen, universitas dan sinta score overall.



## 01 - APA ITU SCRAPING?

- Scraping adalah teknik pengambilan data secara otomatis dari situs web. Proses ini melibatkan penggunaan program untuk mengakses halaman web, mengekstrak informasi yang relevan, dan menyimpannya dalam format yang terstruktur untuk analisis lebih lanjut.

## 02 - WEBSITE SINTA

- Situs SINTA (Science and Technology Index) Kemdikbud adalah platform yang dikembangkan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Indonesia. SINTA menyediakan akses ke data kinerja ilmiah dan teknologi di Indonesia, termasuk publikasi penelitian, sitasi, dan kolaborasi penelitian. Situs ini berfungsi sebagai alat untuk mengukur dan memonitor kinerja peneliti dan institusi pendidikan tinggi di Indonesia, serta untuk meningkatkan kualitas dan produktivitas penelitian nasional.

## 03 - TUJUAN

- Tujuannya untuk mengetahui informasi departemen, universitas maupun score overall dari masing-masing nama peneliti yang tersedia dalam situs sinta sehingga dapat mengevaluasi kontribusi individu peneliti terhadap literatur ilmiah melalui SINTA score overall maupun untuk memonitor tren penelitian, mengidentifikasi potensi kolaborasi, dan membandingkan kinerja antar peneliti serta institusi.

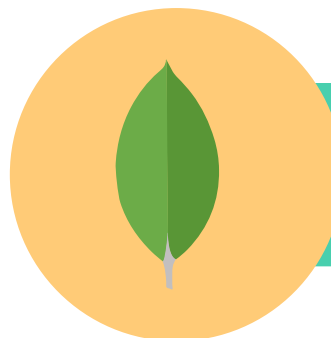


# SCRAPING



## Aplikasi R

- Membuat codingan scraping berdasarkan website sinta : <https://sinta.kemdikbud.go.id/authors?page=> yang kemudian digunakan untuk mengambil informasi beberapa data lalu disambungkan pada Atlas MongoDB.



## MongoDB

- Membuat proyek dan membuat alur penyimpanan database yang berupa atlas collection, atlas DB maupun atlas URL yang kemudian dikaitkan pada GitHub.



## GitHub

- Pada GitHub membuat file scraping dari R dan workflows yang kemudian melakukan pengambilan data melalui workflows dan mengaitkan data pada atlas MongoDB. Kemudian dari atlas MongoDB data di ambil melalui R.

# DATA



IPB University  
— Bogor Indonesia —

Data yang di ambil sebanyak 30 data, mencakup :

- Nama : Nama lengkap dari peneliti yang terdaftar dalam SINTA.
- SINTA ID : Identifikasi unik yang digunakan dalam sistem SINTA untuk mengidentifikasi peneliti.
- Departemen : Unit atau bagian di universitas tempat peneliti terafiliasi atau bekerja.
- Universitas: Institusi pendidikan tinggi di mana peneliti berafiliasi.
- SINTA Score Overall: Skor keseluruhan dari SINTA yang mencerminkan kinerja peneliti berdasarkan berbagai faktor seperti publikasi, sitasi, dan aktivitas ilmiah lainnya.



IPB University  
— Bogor Indonesia —

# ATLAS MONGODB

## scraping2.mds2

STORAGE SIZE: 4KB LOGICAL DATA SIZE: 4.45KB TOTAL DOCUMENTS: 30 INDEXES TOTAL SIZE: 4KB

Find

Indexes

Schema Anti-Patterns 0

Aggregation

Search Indexes

[Generate queries from natural language in Compass](#)

INSERT DOCUMENT

Filter

Type a query: { field: 'value' }

Reset

Apply

[Options](#)

QUERY RESULTS: 1-20 OF MANY

```
_id: ObjectId('6677c0e2111b7755e90b5491')
NAMA : "ASEP BAYU DANI NANDIYANTO"
SINTA_ID : "5974504"
DEPT : "Kimia (S2)"
UNIV : "Universitas Pendidikan Indonesia"
SCORE : "15.992"
```

```
_id: ObjectId('6677c0e2111b7755e90b5491')
```

< PREVIOUS

1-20 of many results

NEXT >



# EKSPLORASI DATA

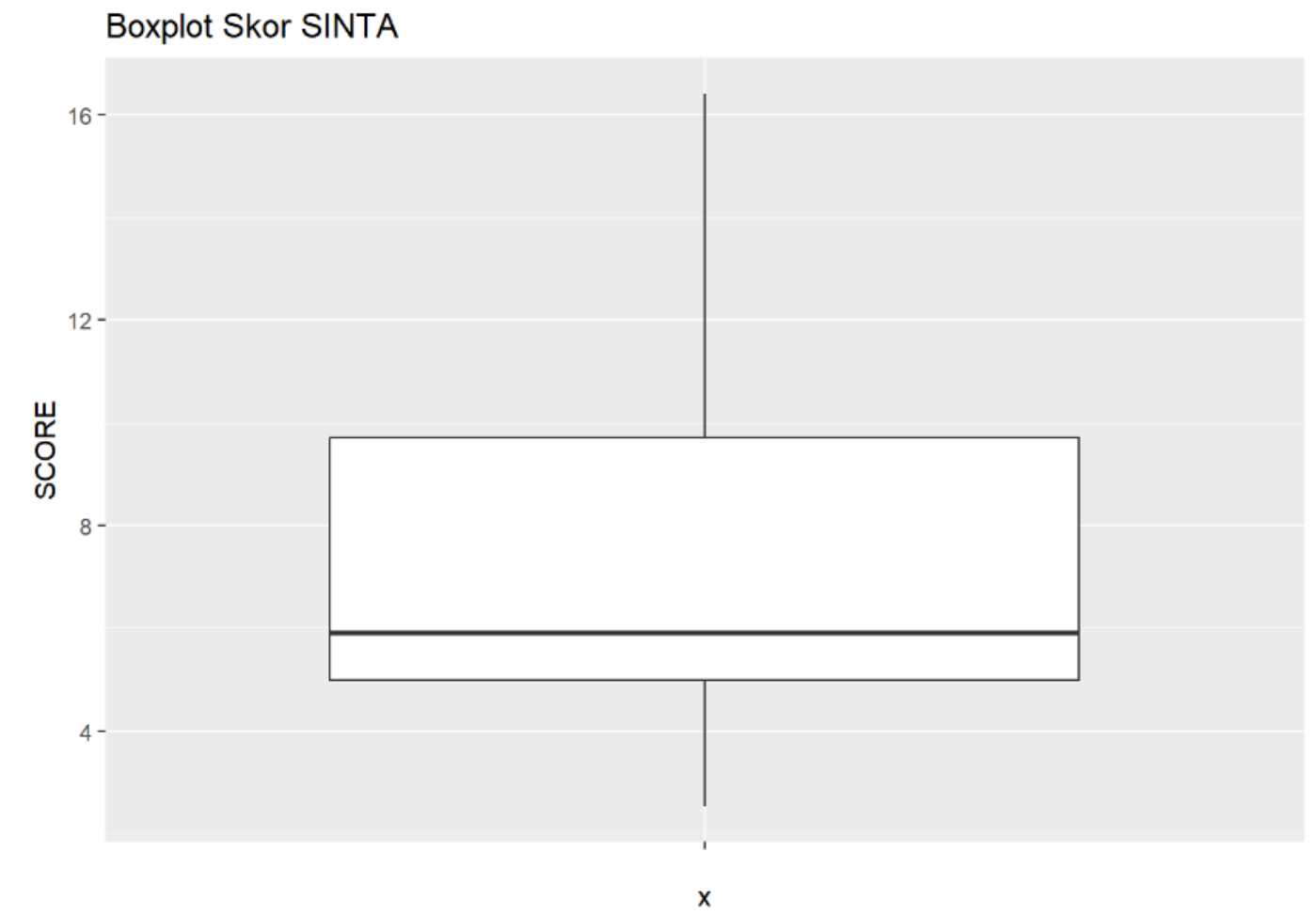


IPB University  
— Bogor Indonesia —

## STATISTIKA DESKRIPTIF DAN BOXPLOT SKOR SINTA

Min.	1st Qu.	Median	Mean	3rd Qu.	Max
2.539	5.000	5.913	7.396	9.704	16.412

- Berdasarkan statistik deskriptif dan boxplot di atas, skor SINTA menunjukkan distribusi yang relatif konsisten. Nilai skor SINTA berkisar dari 2.539 hingga 16.412, dengan median di 5.913. Sebagian besar skor berada dalam rentang 5.000 hingga 9.704, mencakup 50% data tengah.
- Distribusi data sedikit miring ke kanan, ditunjukkan oleh median yang lebih dekat ke kuartil pertama dibandingkan kuartil ketiga. Tidak ada outliers yang terlihat dalam data, yang menunjukkan homogenitas distribusi skor SINTA di antara peneliti. Rata-rata skor adalah 7.396, sedikit lebih tinggi dari median, mengindikasikan beberapa skor yang lebih tinggi mempengaruhi rata-rata keseluruhan.



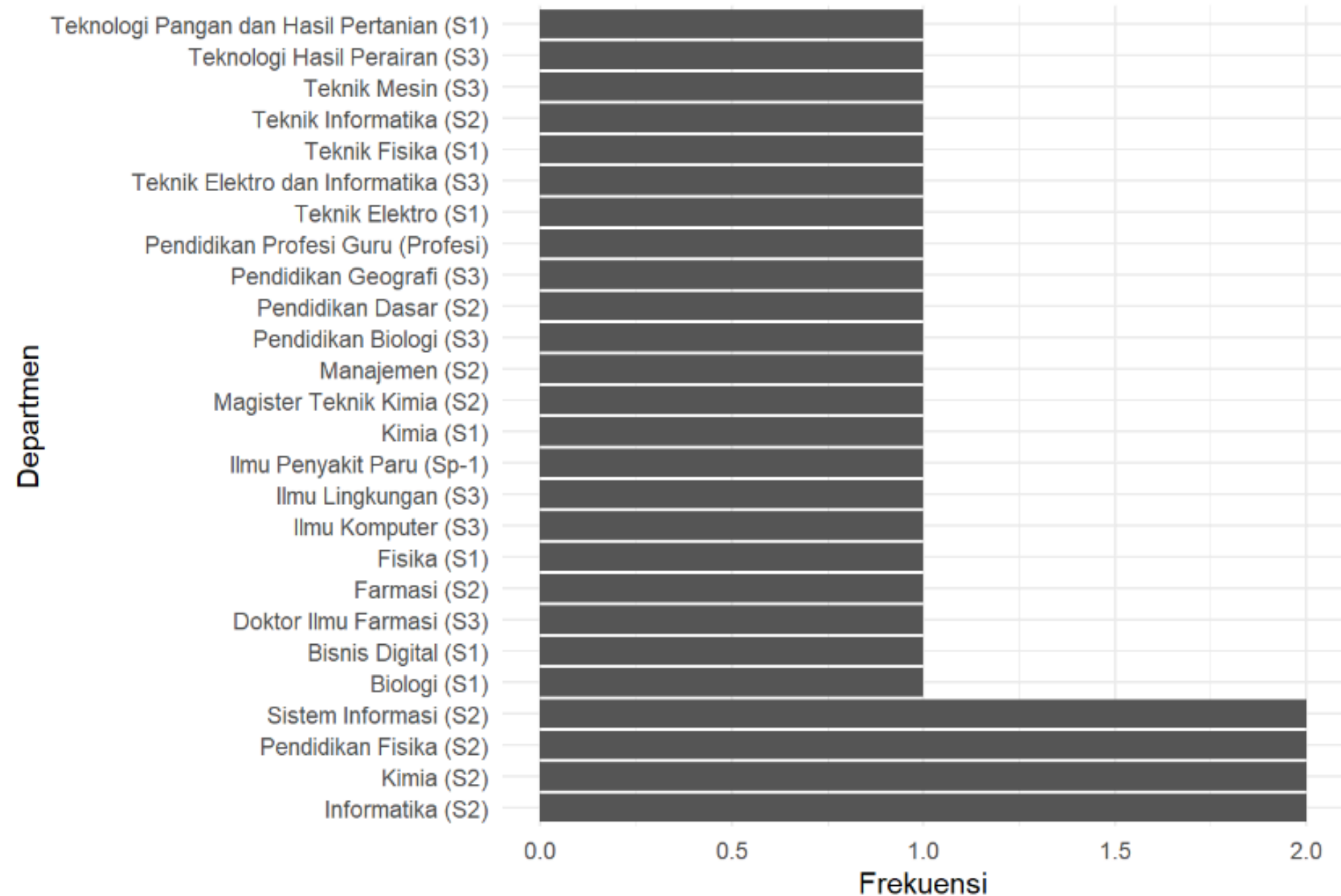


# EKSPLORASI DATA



IPB University  
— Bogor Indonesia —

Jumlah Peneliti dari Departemen

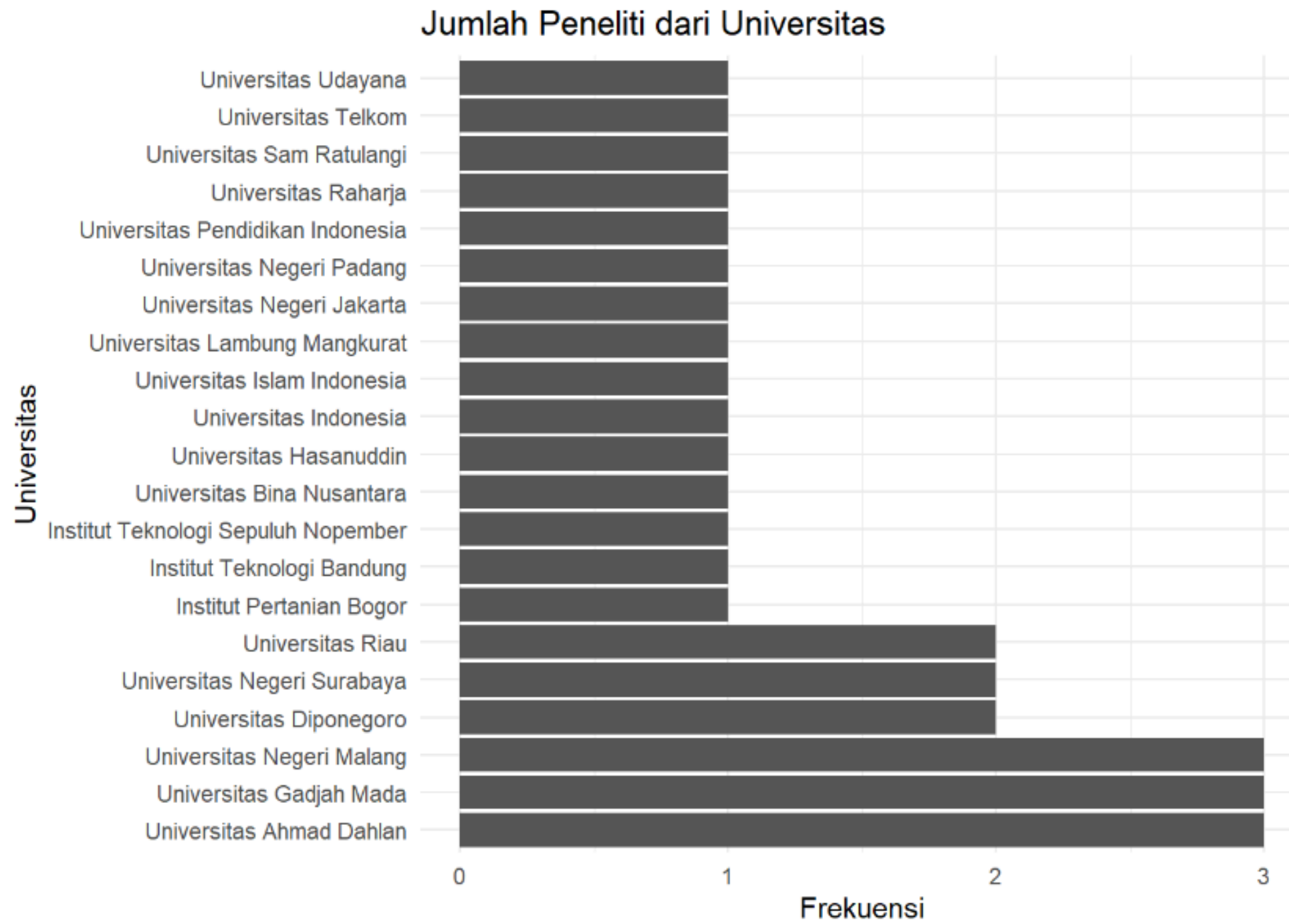


Barplot ini menunjukkan jumlah peneliti di setiap departemen. Departemen dengan jumlah peneliti tertinggi adalah sistem informasi, pendidikan fisika, kimia dan informatika dengan 2 peneliti dan sama-sama memiliki jenjang pendidikan S2. Sebagian besar departemen memiliki 1 peneliti dan grafik ini menunjukkan distribusi yang relatif merata, dengan sedikit variasi di mana satu departemen memiliki sedikit lebih banyak peneliti dibandingkan yang lain.

# EKSPLORASI DATA



IPB University  
— Bogor Indonesia —



Berdasarkan barplot di atas, terlihat bahwa dari 30 data yang diambil setiap peneliti memiliki sebaran latar belakang universitas yang berbeda, terdapat 3 universitas yang memiliki jumlah peneliti terbanyak yaitu memiliki 3 peneliti yang berasal dari Universitas Negeri Malang, Universitas Gajah Mada, dan Universitas Ahmad Dahlan.

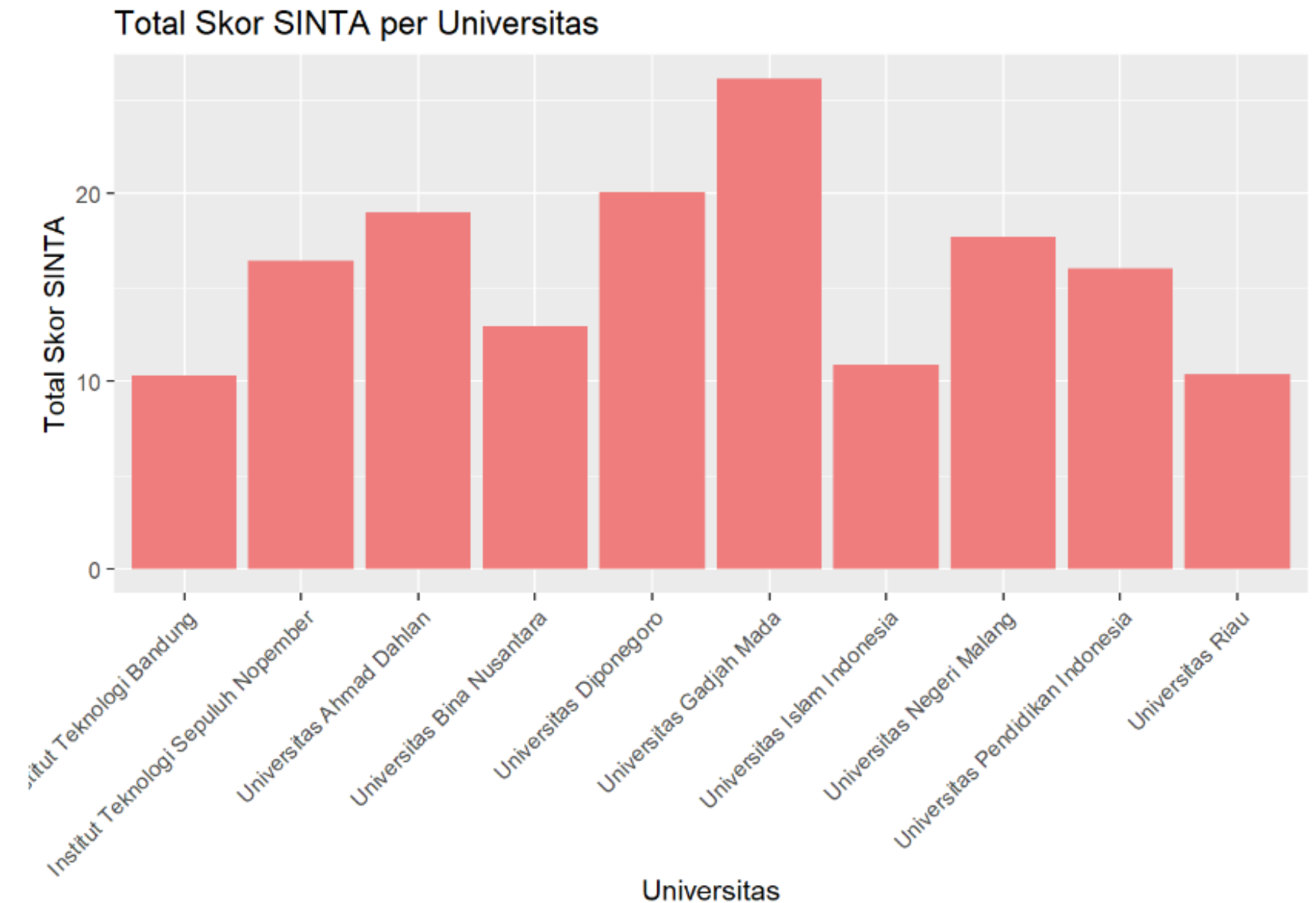
# EKSPLORASI DATA



IPB University  
— Bogor Indonesia —

UNIV <fctr>	total_score <dbl>
Universitas Gadjah Mada	26.111
Universitas Diponegoro	20.085
Universitas Ahmad Dahlan	18.986
Universitas Negeri Malang	17.726
Institut Teknologi Sepuluh Nopember	16.412
Universitas Pendidikan Indonesia	15.992
Universitas Bina Nusantara	12.936
Universitas Islam Indonesia	10.890
Universitas Riau	10.363
Institut Teknologi Bandung	10.298

1-10 of 10 rows



Berdasarkan barplot dan data total score dari sinta score overall, terlihat bahwa Universitas Gajah Mada memiliki total score tertinggi yaitu 26.111 dari 3 peneliti, lalu di ikuti Universitas Diponegoro dan Universitas Ahmad Dahlan dengan rentang nilai 10 total score tertinggi nya adalah 10.298 sampai 26.111

# EKSPLORASI DATA



IPB University  
— Bogor Indonesia —

## TOP 10 DATA TERTINGGI BERDASARKAN SKOR SINTA

Description: df [10 × 5]

NAMA <fctr>	SINTA_ID <dbl>	DEPT <fctr>	UNIV <fctr>	SCORE <dbl>
RIYANARTO SARNO	29555	Ilmu Komputer (S3)	Institut Teknologi Sepuluh Nopember	16.412
ASEP BAYU DANI NANDIYANTO	5974504	Kimia (S2)	Universitas Pendidikan Indonesia	15.992
ABDUL ROHMAN	42423	Doktor Ilmu Farmasi (S3)	Universitas Gadjah Mada	14.705
BENS PARDAMEAN	6043909	Teknik Informatika (S2)	Universitas Bina Nusantara	12.936
IS FATIMAH	5998917	Kimia (S2)	Universitas Islam Indonesia	10.890
HADIYANTO	5049	Sistem Informasi (S2)	Universitas Diponegoro	10.707
BRIAN YULIARTO	260433	Teknik Fisika (S1)	Institut Teknologi Bandung	10.298
Tole Sutikno	23026	Informatika (S2)	Universitas Ahmad Dahlan	9.812
JAMARI	6023233	Teknik Mesin (S3)	Universitas Diponegoro	9.378
SITI ZUBAIDAH	5988462	Pendidikan Biologi (S3)	Universitas Negeri Malang	8.158

1-10 of 10 rows

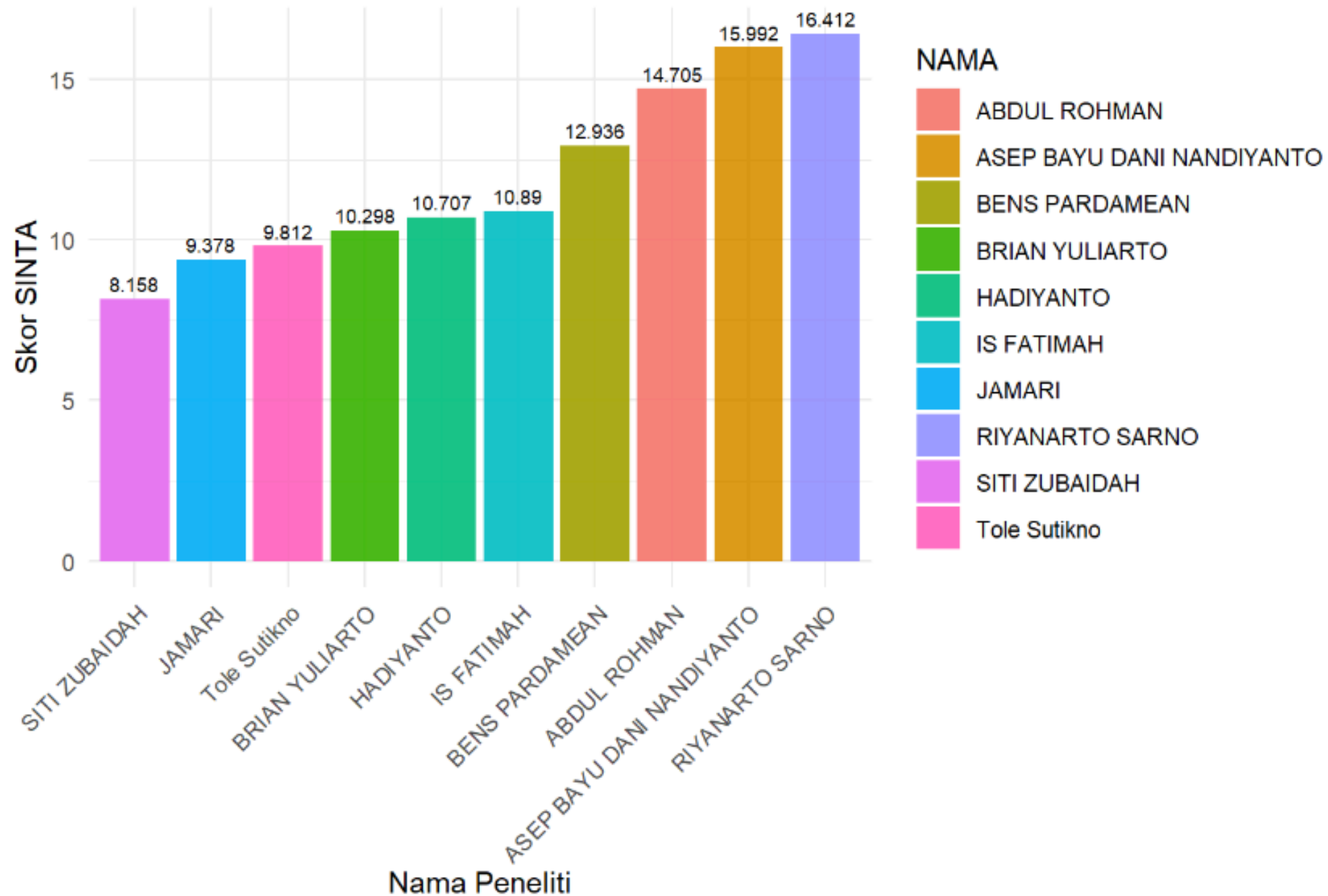


# EKSPLORASI DATA



IPB University  
— Bogor Indonesia —

Skor SINTA untuk 10 Peneliti Tertinggi



Berdasarkan barplot dan data sinta score overall, terlihat bahwa sinta score overall berkisar dari 8.158 hingga 16.412, dengan distribusi yang menunjukkan perbedaan signifikan antara peneliti dengan skor terendah dan tertinggi. Score tertinggi di peroleh oleh Riyanarto Sarno dengan sinta score overall 16.412 berasal dari Institut Teknologi Sepuluh Nopember dengan Departement Ilmu Komputer (S3).



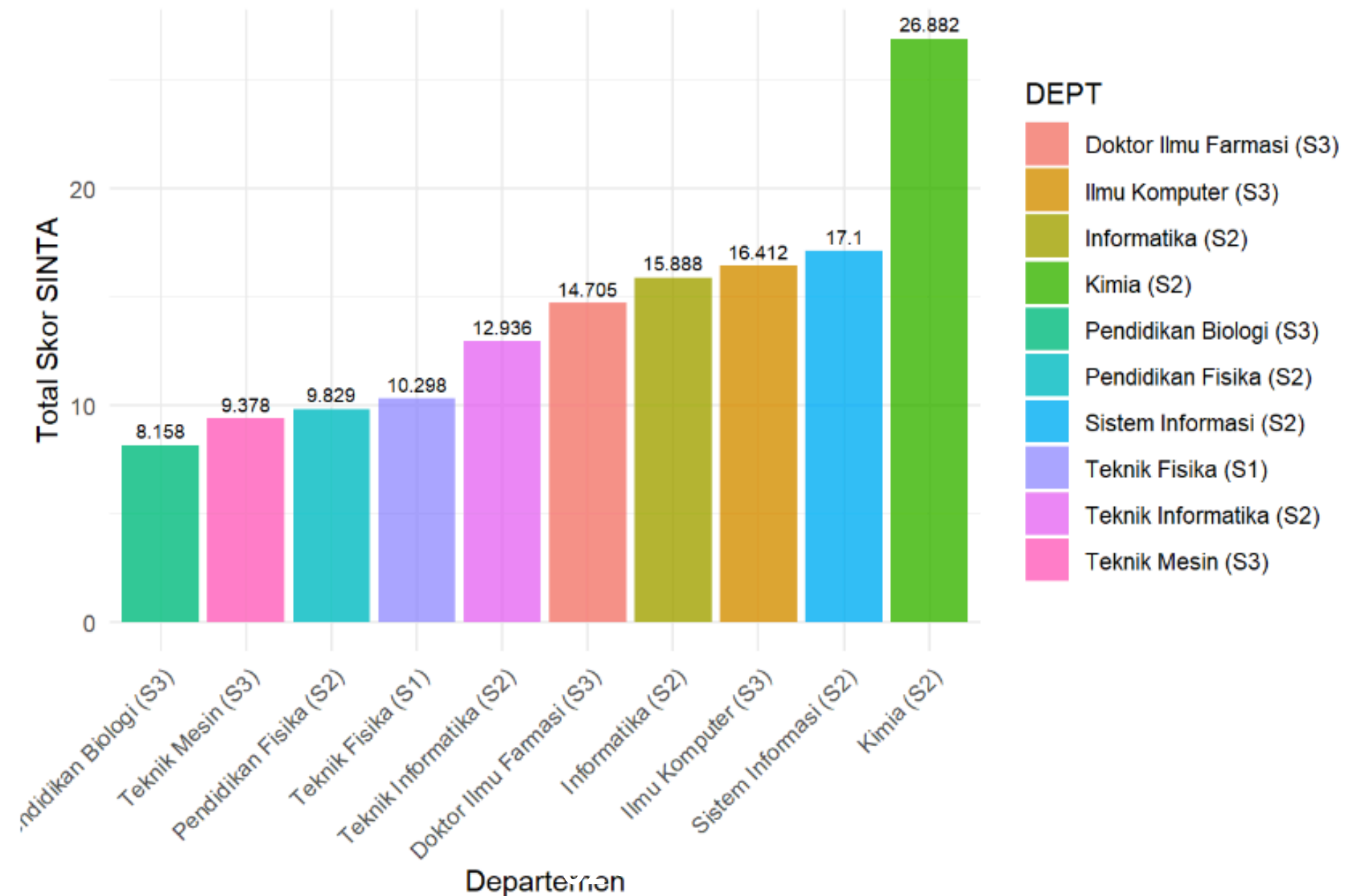
# EKSPLORASI DATA



IPB University  
— Bogor Indonesia —

Berdasarkan barplot dan data total score dari sinta score overall terhadap departemen, terlihat bahwa score tertinggi di peroleh dari departement Kimia jenjang S2 dengan sinta score overall 26.882, lalu urutan kedua dari departement Sistem Informasi Jenjang S2 dan urutan ketiga dari departement Ilmu Komputer jenjang S3 dengan rentang total score dari 8.158 sampai 26.882

Total Skor SINTA per Departemen (Top 10)



# EKSPLORASI DATA

Berdasarkan word cloud di atas terlihat bahwa departement Informatika (S2), Kimia (S2), Pendidikan Fisika (S2), dan Sistem Informasi (S2) muncul paling besar dalam word cloud dan memiliki warna yang berbeda dari yang lainnya, artinya keempat departemen tersebut memiliki jumlah peneliti yang paling banyak dibandingkan dengan departement lain.





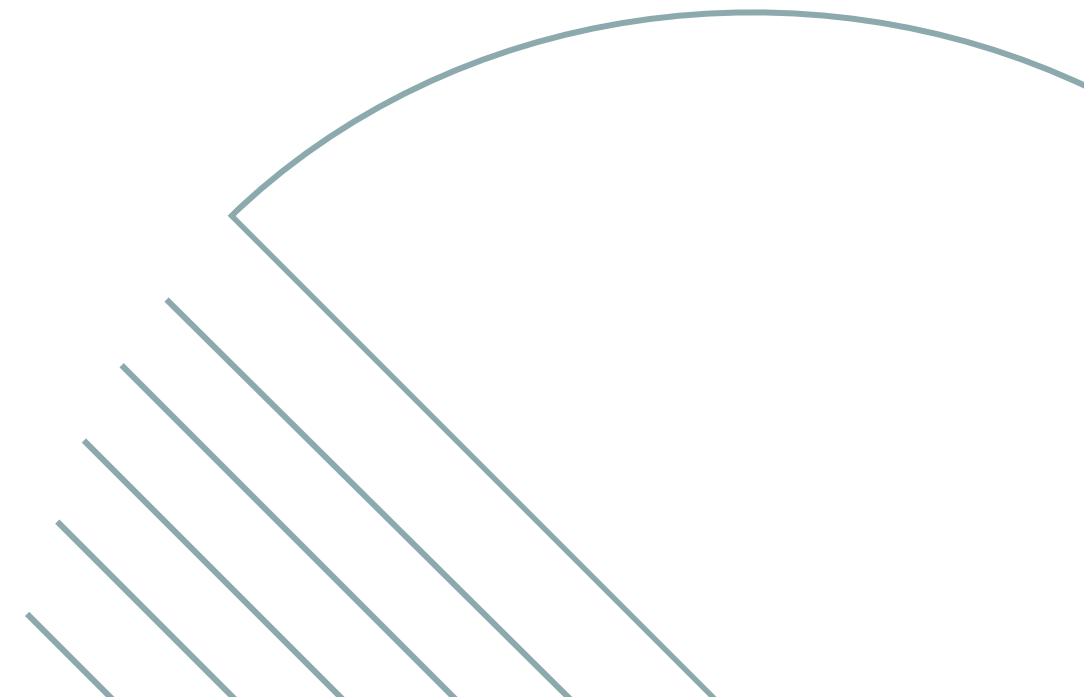
# KESIMPULAN



IPB University  
— Bogor Indonesia —



Dari hasil scraping situs SINTA, dapat disimpulkan bahwa terdapat variasi yang signifikan dalam skor SINTA antar universitas dan departemen. Universitas dengan skor SINTA tertinggi cenderung memiliki reputasi yang lebih baik dalam hal kontribusi akademik. Analisis departemen menggunakan word cloud mengungkapkan fokus penelitian yang dominan di setiap institusi, memberikan wawasan tentang tren penelitian di tingkat departemen. Informasi ini dapat menjadi panduan strategis bagi institusi dalam mengarahkan fokus penelitian dan mengoptimalkan dampak akademik mereka.





# THANK YOU

[https://github.com/yunnamentari/scraping\\_mds](https://github.com/yunnamentari/scraping_mds)