

# TRAINING MELEK DATA

## INTRODUCTION

Ikang Fadhli

Nutrifood Indonesia

25 May 2022

- 1 *OUTLINE TRAINING*
- 2 *INTRODUCTION TO DATA*
- 3 *KNOWLEDGE ABOUT DIFFERENT DATA TYPES AND FORMATS*
- 4 *BASIC KNOWLEDGE AND UNDERSTANDING DATA*
- 5 *APPLICATIONS OF DATA*

## Section 1

# *OUTLINE TRAINING*

# OUTLINE TRAINING

Materi pada *training* ini mencakup:

- *Introduction to data*
- *Knowledge about different data types and formats*
- *Basic knowledge and understanding data*
- *Understanding of relevance and applications of data*

## Section 2

# *INTRODUCTION TO DATA*

## Subsection 1

### *Apa Sih Data Itu?*

# Definisi Data

Salah satu definisi data yang sering saya pakai adalah:

## Definisi

Data adalah representasi faktual dari observasi (pengamatan).

Data bisa dikelompokkan sesuai dengan tipe dan karakteristiknya. Dengan memahami tipe data yang kita hadapi, kita akan lebih mudah dalam menentukan metode analisa yang tepat.

# Data itu Seperti Apa?

Pada 10 tahun yang lalu, saat kita mendengar istilah **data**. Hal yang terlintas pertama kali adalah tabel dalam program Microsoft Excel.

A		B	C	D	E
Column1	Month	DATE	ITEM		
1	HEAVENLY BLUSH, YOGURT DRINK PEACH TPK 200mL 42522	JUN 16	1	HEAVENLY BLUSH, YOGURT DRINK PEACH TPK 200mL	
2	L-MEN, SUSU BUBUK GAIN MASS FOR MEN CHOCOLATE BOX 225g 42522	JUN 16	1	L-MEN, SUSU BUBUK GAIN MASS FOR MEN CHOCOLATE BOX 225g	
3	TROPICANA SLIM, SUSU BUBUK LOW FAT MILK VANILLA BOX 180g 42522	JUN 16	1	TROPICANA SLIM, SUSU BUBUK LOW FAT MILK VANILLA BOX 180g	
4	WRP, DIET DRINK STAY SLIM CHOCO HAZELNUT BOX 250g 42522	JUN 16	1	WRP, DIET DRINK STAY SLIM CHOCO HAZELNUT BOX 250g	
5	TROPICANA SLIM, SWEETENER PK 50x2.5 g 42522	JUN 16	1	TROPICANA SLIM, SWEETENER PK 50x2.5 g	
6	TROPICANA SLIM, SWEETENER NO CALORIE DIABETICS BOX 50x2g 42522	JUN 16	1	TROPICANA SLIM, SWEETENER NO CALORIE DIABETICS BOX 50x2g	
7	L-MEN, SUSU CAIR UHT HI-PROTEIN 2 GO CHOCOLATE BOX 2X200mL 42522	JUN 16	1	L-MEN, SUSU CAIR UHT HI-PROTEIN 2 GO CHOCOLATE BOX 2X200mL	
8	NUTRISARI, INSTANT DRINK JERUK PERAS PK 5X14g 42522	JUN 16	1	NUTRISARI, INSTANT DRINK JERUK PERAS PK 5X14g	
9	WRP, SUSU CAIR ON THE GO COFFEE TPK 200mL 42522	JUN 16	1	WRP, SUSU CAIR ON THE GO COFFEE TPK 200mL	
10	NUTRISARI, JUICE DRINK FRUIT&VEGETABLE TPK 200mL 42522	JUN 16	1	NUTRISARI, JUICE DRINK FRUIT&VEGETABLE TPK 200mL	
11	HILO SCHOOL, SUSU BUBUK HI-CALCIUM CHOCOLATE BOX 250g 42522	JUN 16	1	HILO SCHOOL, SUSU BUBUK HI-CALCIUM CHOCOLATE BOX 250g	
12	L-MEN, WATER JERUK BTL 350mL 42522	JUN 16	1	L-MEN, WATER JERUK BTL 350mL	
13	HILO SCHOOL, SUSU BUBUK HI-CALCIUM VANILLA BOX 250g 42522	JUN 16	1	HILO SCHOOL, SUSU BUBUK HI-CALCIUM VANILLA BOX 250g	
14	TROPICANA SLIM, SWEETENER LOW CALORIE BOX 250g 42522	JUN 16	1	TROPICANA SLIM, SWEETENER LOW CALORIE BOX 250g	
15	NUTRISARI, INSTANT DRINK AMARANTH SWEET ORANGE PK 10X14g 42522	JUN 16	1	NUTRISARI, INSTANT DRINK AMARANTH SWEET ORANGE PK 10X14g	
16	TROPICANA SLIM, MILK HI LO CALCIUM FAT CHOCOLATE BOX 250g 42522	JUN 16	1	TROPICANA SLIM, MILK HI LO CALCIUM FAT CHOCOLATE BOX 250g	
17	HILO TEEN, SUSU CAIR UHT CHOCOLATE TPK 200mL 42522	JUN 16	1	HILO TEEN, SUSU CAIR UHT CHOCOLATE TPK 200mL	
18	HILO TEEN, SUSU BUBUK HI-CALCIUM COOLAT BOX 500g 42522	JUN 16	1	HILO TEEN, SUSU BUBUK HI-CALCIUM COOLAT BOX 500g	
19	WRP, SUSU CAIR ON THE GO CHOCOLATE TPK 200mL 42522	JUN 16	1	WRP, SUSU CAIR ON THE GO CHOCOLATE TPK 200mL	
20	NUTRISARI, INSTANT DRINK FLORIDA ORANGE PK 10X14g 42522	JUN 16	1	NUTRISARI, INSTANT DRINK FLORIDA ORANGE PK 10X14g	
21	TROPICANA SLIM, MILK HI LO CALCIUM TEEN VANILLA CARAMEL BOX 250g 42522	JUN 16	1	TROPICANA SLIM, MILK HI LO CALCIUM TEEN VANILLA CARAMEL BOX 250g	
22	HILO, SUSU BUBUK HI-CALCIUM BELGIAN CHOCO PK 10x30g 42522	JUN 16	1	HILO, SUSU BUBUK HI-CALCIUM BELGIAN CHOCO PK 10x30g	
23	HEAVENLY BLUSH, YOGURT DRINK BLACKCURRANT TPK 200mL 42522	JUN 16	1	HEAVENLY BLUSH, YOGURT DRINK BLACKCURRANT TPK 200mL	
24	NUTRISARI, MINUMAN W/DAIRY SAIGUR BOX 4X15g 42522	JUN 16	1	NUTRISARI, MINUMAN W/DAIRY SAIGUR BOX 4X15g	
25	HEAVENLY BLUSH, YOGURT DRINK BANANA BERRIES BROCCOLI TPK 200mL 42522	JUN 16	1	HEAVENLY BLUSH, YOGURT DRINK BANANA BERRIES BROCCOLI TPK 200mL	
26	HEAVENLY BLUSH, YOGURT STRAWBERRY TPK 200mL 42522	JUN 16	1	HEAVENLY BLUSH, YOGURT STRAWBERRY TPK 200mL	
27	HEAVENLY BLUSH, YOGURT DRINK MANGO CARROT TPK 200mL 42522	JUN 16	1	HEAVENLY BLUSH, YOGURT DRINK MANGO CARROT TPK 200mL	
28	HEAVENLY BLUSH, GREEK YOGURT HONEY CUP 100g 42522	JUN 16	1	HEAVENLY BLUSH, GREEK YOGURT HONEY CUP 100g	
29	HEAVENLY BLUSH, GREEK YOGURT PEACH CUP 100g 42522	JUN 16	1	HEAVENLY BLUSH, GREEK YOGURT PEACH CUP 100g	
30	TROPICANA SLIM, ALEGRON COOKIES PK 100g 42522	JUN 16	1	TROPICANA SLIM, ALEGRON COOKIES PK 100g	
31	HEAVENLY BLUSH, GREEK YOGURT STRAWBERRY CUP 100g 42522	JUN 16	1	HEAVENLY BLUSH, GREEK YOGURT STRAWBERRY CUP 100g	
32	TROPICANA SLIM, SUGAR FREE DRINK I.TL.CAFFE LATTE BOX 10X14g 42522	JUN 16	1	TROPICANA SLIM, SUGAR FREE DRINK I.TL.CAFFE LATTE BOX 10X14g	
33	L-MEN, SUSU BUBUK GAIN MASS FOR MEN CHOCOLATE BOX 225g 42522	JUN 16	1	L-MEN, SUSU BUBUK GAIN MASS FOR MEN CHOCOLATE BOX 225g	
34	TROPICANA SLIM, KECAP DIET MANIS BTL 200mL 42522	JUN 16	1	TROPICANA SLIM, KECAP DIET MANIS BTL 200mL	
35	WRP, DIET DRINK CHOCOLATE BOX 12X25g 42522	JUN 16	1	WRP, DIET DRINK CHOCOLATE BOX 12X25g	
36	TROPICANA SLIM, KECAP DIET ASIN BTL 200mL 42522	JUN 16	1	TROPICANA SLIM, KECAP DIET ASIN BTL 200mL	
37	HEAVENLY BLUSH, YOGURT DRINK PEACH TPK 200mL 42522	JUN 16	2	HEAVENLY BLUSH, YOGURT DRINK PEACH TPK 200mL	
38	L-MEN, SUSU BUBUK GAIN MASS FOR MEN CHOCOLATE BOX 225g 42522	JUN 16	2	L-MEN, SUSU BUBUK GAIN MASS FOR MEN CHOCOLATE BOX 225g	
39	TROPICANA SLIM, SUSU BUBUK LOW FAT MILK VANILLA BOX 180g 42522	JUN 16	2	TROPICANA SLIM, SUSU BUBUK LOW FAT MILK VANILLA BOX 180g	
40	WRP, DIET DRINK STAY SLIM CHOCO HAZELNUT BOX 250g 42522	JUN 16	2	WRP, DIET DRINK STAY SLIM CHOCO HAZELNUT BOX 250g	
41	TROPICANA SLIM, SWEETENER PK 50x2.5 g 42522	JUN 16	2	TROPICANA SLIM, SWEETENER PK 50x2.5 g	
42	TROPICANA SLIM, SWEETENER NO CALORIE DIABETICS BOX 50x2g 42522	JUN 16	2	TROPICANA SLIM, SWEETENER NO CALORIE DIABETICS BOX 50x2g	
43	L-MEN, SUSU CAIR UHT HI-PROTEIN 2 GO CHOCOLATE BOX 2X200mL 42522	JUN 16	2	L-MEN, SUSU CAIR UHT HI-PROTEIN 2 GO CHOCOLATE BOX 2X200mL	
44	NUTRISARI, INSTANT DRINK JERUK PERAS PK 5X14g 42522	JUN 16	2	NUTRISARI, INSTANT DRINK JERUK PERAS PK 5X14g	
45	WRP, SUSU CAIR ON THE GO COFFEE TPK 200mL 42522	JUN 16	2	WRP, SUSU CAIR ON THE GO COFFEE TPK 200mL	

Figure 1: Contoh Data Ms. Excel



# Data Excel

Bentuk data seperti itu adalah bentuk data “termudah” untuk bisa dianalisa. Setidaknya kita bisa melakukan analisa seperti *pivoting*, *pareto*, dan lainnya.

Namun bagaimana dengan data dengan bentuk lainnya?

# Data Bentuk Lainnya

Pada era digital ini, data sudah mulai bervariasi bentuknya seperti:

- 1 *Text,*
- 2 *Figures,*
- 3 *Videos,*
- 4 *Geolocation,*
- 5 *dll.*

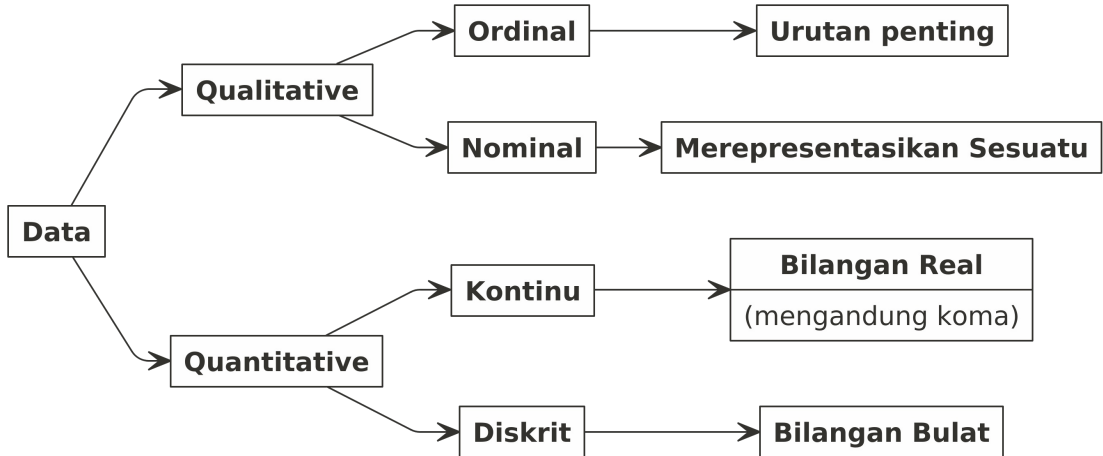
## Section 3

# ***KNOWLEDGE ABOUT DIFFERENT DATA TYPES AND FORMATS***

## Subsection 1

### Pengelompokan Data Berdasarkan Tipe (Statistik)

# Pengelompokkan Data Berdasarkan Tipe (Statistik)



# Data Kualitatif

## Definisi

Data Kualitatif adalah data yang tidak bisa dilakukan operasi aritmatika (penjumlahan, pengurangan, pembagian, dan perkalian). Data seperti ini bisa juga disebut sebagai data kategorik.

- Nominal; Representasi dari sesuatu.
  - Contoh: data seperti *gender*. Misalkan angka 1 saya tulis sebagai representasi dari pria dan 2 sebagai wanita.
- Ordinal; Urutan dari data menjadi penting.
  - Contoh: skala *likert*, misalkan angka 1 - 6 sebagai representasi dari tingkat kesukaan atau kesetujuan (sangat suka sampai sangat tidak suka).

# Data Kuantitatif

## Definisi

Data Kuantitatif adalah data yang bisa dilakukan operasi aritmatika (penjumlahan, pengurangan, pembagian, dan perkalian). Data seperti ini, kita akan sebut sebagai data numerik.

- Diskrit; bilangan bulat (*integer*).
  - Contoh: banyaknya anak, banyaknya karyawan, dll.
- Kontinu; bilangan *real* (mengandung koma).
  - Contoh: tinggi badan, berat badan, dll.

## Contoh Penerapan

Misalkan dalam satu kelas *webinar* berisi 10 orang peserta dengan *gender*:

- 1 7 orang wanita
- 2 3 orang pria

Apakah pernyataan berikut ini benar / salah?

***Rata-rata peserta webinar tersebut adalah wanita.***



## Contoh Penerapan

Pernyataan tersebut **SALAH!**

## Contoh Penerapan

### Kenapa?

Bagaimana cara menghitung rata-rata?

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

**Sekarang kita hitung data kita:**

Misalkan saya tulis pria sebagai 1 dan wanita sebagai 2. Maka didapatkan:

$$\sum_{i=1}^n x_i = 1 + 1 + 1 + 2 + 2 + 2 + 2 + 2 + 2 + 2 = 17$$

$$n = 10$$

$$\bar{x} = \frac{17}{10} = 1.7$$

## Contoh Penerapan

Apa arti 1.7?

*Pria yang terlalu kewanitaan?*

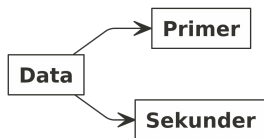
*Wanita yang agak kelaki-lakian?*

Pernyataan yang **Tepat** adalah **Mayoritas** peserta *webinar* adalah wanita.

## Subsection 2

### Pengelompokan Data Berdasarkan Sumber

# Pengelompokan Data Berdasarkan Sumber



## Definisi

Data primer adalah data yang diambil dari sumber pertama.

Data sekunder adalah data yang berasal dari sumber yang pernah diolah oleh orang lain.

# Contoh Data Primer dan Sekunder

## Data Primer

- ➊ Data yang kita himpun sendiri dari hasil interview menggunakan quesioner (*survey*),
- ➋ Data yang kita ambil dari mesin,
- ➌ Data hasil *web scrape*, dll.

## Data Sekunder

- ➊ Data kependudukan hasil sensus BPS,
- ➋ Laporan absensi karyawan,
- ➌ Data Covid-19 yang tersedia di *website* resmi pemerintah, dll.

# Masalah Seputar Data Primer dan Sekunder

## Pertanyaannya:

*Kapan kita memakai **data primer**? Kapan kita memakai **data sekunder**?*

## Guidance untuk menjawab pertanyaan:



## Penjelasan *Guidance*

Prinsipnya adalah:

*Kita akan menggunakan data primer saat data sekunder tidak ada!*

Sejujurnya mencari data primer itu relatif sulit. Setidaknya kita membutuhkan waktu, tenaga, dan biaya untuk mencari data langsung dari sumbernya. Contoh: jika saya ingin mencari tahu berapa banyak orang yang teridentifikasi **COVID-19** di suatu kota. Alih-alih saya datang ke semua RS yang ada di kota tersebut, saya cukup cek saja *website* resmi penanganan **Covid-19** pemerintah.

Jadi, jika data sekundernya sudah tersedia kita bisa mempertimbangkan untuk memakai data tersebut daripada mengambil data primer. **TAPI** jika ternyata karakteristik data yang kita mau cari tersebut **sangat dinamis dan cepat berubah** ATAU ada **perbedaan kondisi, situasi atau limitasi** maka kita harus **mempertimbangkan untuk mencari data primer dan tidak menggunakan data sekunder**.



## Subsection 3

### Pengelompokan Data Berdasarkan Tipe (Data Sains)

# Pengelompokan Data Berdasarkan Tipe (Data Sains)

Di dalam dunia data sains, ada beberapa tipe data yang sering digunakan. Secara hierarki, bisa diurutkan sebagai berikut:

`character > numeric > integer > logical`

# Pengelompokan Data Berdasarkan Tipe (Data Sains)

Berikut penjelasannya:

- ➊ **character**: merupakan tipe data berupa karakter atau *string*. Semua data bisa dilihat sebagai *character*. Oleh karena itu, secara hierarki tipe data ini ditempatkan di urutan paling atas. Namun, data tipe ini tidak bisa dilakukan operasi aritmatika.
- ➋ **numeric**: merupakan tipe data angka berupa bilangan *real*. Tipe data ini mirip dengan data numerik di poin sebelumnya.
- ➌ **integer**: merupakan tipe data angka berupa bilangan bulat. Sekilas mirip dengan tipe data diskrit di poin sebelumnya. Namun di beberapa kondisi, tipe data ini bisa dijadikan data **kategorik** sehingga kita bisa sebut tipenya menjadi *factor*.
- ➍ **logical**: merupakan tipe data *boolean*. Hanya berisi *TRUE* atau *FALSE*. Tipe data ini sangat berguna saat kita melakukan *if conditional*, *looping*, atau membuat *regex* (*regular expression*).

## Subsection 4

### Data di Era Digital

# Data di Era Digital

Era digital seperti sekarang membuat arus penambahan data sangatlah kencang. Akibatnya muncul tiga masalah utama terkait data:

- ❶ Di mana mencari datanya? Bagaimana cara mengambilnya?
- ❷ Tidak semua data yang ada bersih.
- ❸ Bagaimana mengesktrak *insights* dari data?

## Section 4

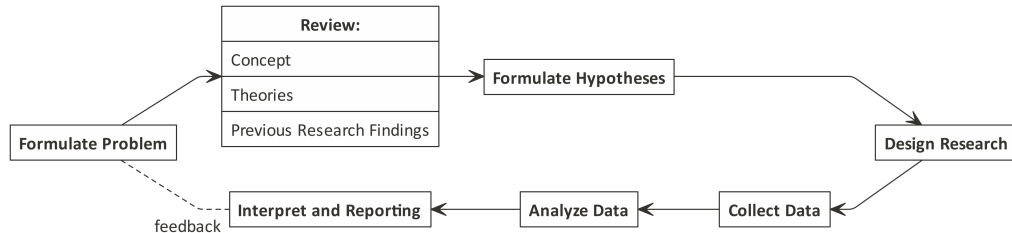
# BASIC KNOWLEDGE AND UNDERSTANDING DATA

## Subsection 1

### *Research Workflow*

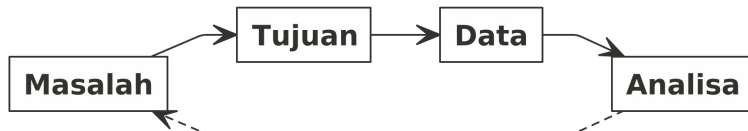
## Proper Steps

Dalam setiap riset atau penelitian, alur pengerjaannya relatif sama seperti ini:





# Hubungan Masalah dengan Data



## Relevansi Data

Grafik di atas memberikan suatu arahan bahwa data yang dikumpulkan dan dianalisa harus sesuai dengan masalah yang dihadapi.

## Subsection 2

### Kualitas dari Suatu Data

# Kualitas dari Suatu Data

## Pertanyaannya:

*Data saya sudah bagus belum ya?*

Ada beberapa parameter yang bisa digunakan untuk menjawab pertanyaan tersebut:

- 1 Kualitas dari suatu data diukur dari **seberapa bisa data tersebut memenuhi kebutuhan analisis**. Setelah itu baru dicek secara statistik: *blank data* atau *extreme values*.
- 2 Data yang bagus biasanya berasal dari *random sampling*.
- 3 Untuk data yang diambil secara berulang-ulang (harian, mingguan, bulanan, tahunan, atau lainnya), pastikan bahwa cara pengambilan data harus sama.

## Subsection 3

### *Importance of Disclaimer*

# Importance of Disclaimer

## Pertanyaannya:

*Apa yang menyebabkan hasil survey / study / penelitian / analisa data tidak sesuai dengan kondisi real?*

Ingat bahwa setiap penelitian memiliki **limitasi** dan **asumsi**.

- 1 Perhatikan dengan seksama kondisi yang dilalui saat melakukan penelitian. Apakah ada perbedaan dengan kondisi *real*?
- 2 Perhatikan cara pengambilan data. Apakah ada *bias* yang kita lakukan saat mengambil data?

# Contoh

**Misalkan:**

Survey mengenai makanan favorit orang Indonesia tapi survey dilakukan di RM Padang.

**Pertanyaannya:**

Kira-kira makanan apa yang menjadi favorit orang Indonesia tersebut?

## Section 5

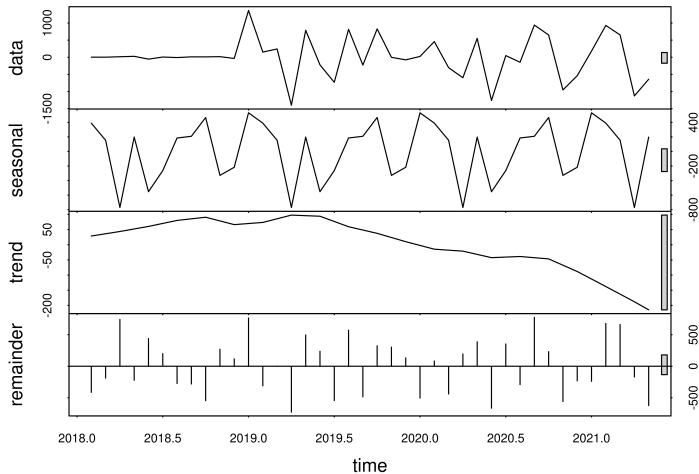
# *APPLICATIONS OF DATA*

## Subsection 1

### Aplikasi Data di *Sales*



# Aplikasi Data di Sales



Tim *sales* memiliki data omset bulanan suatu produk tertentu. Dari data tersebut, bisa kita analisa tren penjualan, efek *seasonal*, efek promosi, pengelompokkan *customer*, dll.

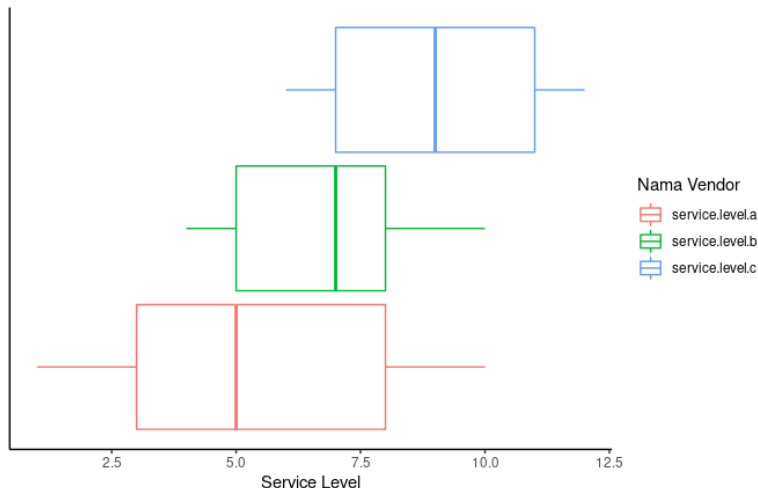
## Subsection 2

### Aplikasi Data di *Procurement*

# Aplikasi Data di *Procurement*

Box Plot dari Masing-Masing Vendor

Data Tiga Vendor



Tim procurement hendak mengambil keputusan strategic terkait 3 vendor yang selama ini digunakan. Mereka mencoba service level antara ketiganya. Siapa yang terbaik? Apakah ketiganya masih sesuai dengan spek minimal yang dipersyaratkan?