

Tugas 2 – Analisis dan Visualisasi Data-MSIM4310	
Nama	Pepy Ilhadi Beltra (048347842)
Prodi	Sistem Informasi

1. Dalam konsep transformasi data menggunakan bahasa R, paket `dplyr` adalah salah satu alat yang sangat populer karena menyediakan fungsi-fungsi yang efisien dan intuitif untuk manipulasi data, terutama pada data frame. Berikut adalah beberapa fungsi transformasi utama dalam `dplyr` beserta penjelasannya:

- 1. `mutate()`
Fungsi ini digunakan untuk menambahkan kolom baru atau memodifikasi kolom yang sudah ada dalam data frame berdasarkan operasi pada kolom lain. Ideal digunakan untuk perhitungan baru atau transformasi data.
Contoh: `mutate(data, total = price * quantity)` menambahkan kolom `total` yang merupakan hasil perkalian kolom `price` dan `quantity`.
Ideal digunakan untuk perhitungan baru atau transformasi data.
- 2. `select()`
Fungsi ini memungkinkan pemilihan kolom tertentu dari data frame untuk dipertahankan, atau membuang kolom yang tidak diinginkan.
Contoh: `select(data, name, age)` hanya menyimpan kolom `name` dan `age`.
Kegunaan dalam Membantu menyederhanakan data dengan memfilter kolom yang relevan.
- 3. `filter()`
Digunakan untuk menyaring baris dalam data frame berdasarkan kondisi tertentu.
Contoh: `filter(data, age > 18)` menyimpan hanya baris di mana nilai `age` lebih dari 18.
Fungsi ini berguna untuk subset data berdasarkan kriteria spesifik.
- 4. `arrange()`
Mengurutkan baris dalam data frame berdasarkan satu atau lebih kolom, baik secara ascending (meningkat) maupun descending (menurun).
Contoh: `arrange(data, desc(age))` mengurutkan data berdasarkan kolom `age` dari terbesar ke terkecil.
Fungsi ini digunakan untuk Membantu mengorganisasi data untuk analisis atau presentasi.
- 5. `summarise()` / `summarize()`
Mengaggregasi data dengan menghitung ringkasan statistik (misalnya, rata-rata, jumlah, maksimum) berdasarkan kelompok atau keseluruhan data.
Contoh: `summarise(data, avg_price = mean(price))` menghitung rata-rata nilai kolom `price`.
Fungsi ini Digunakan untuk analisis data agregat, sering dikombinasikan dengan `group_by()`.
- 6. `group_by()`
Membagi data menjadi kelompok berdasarkan satu atau lebih variabel, biasanya digunakan bersama `summarise()` untuk analisis per kelompok.

Contoh: ``group_by(data, category) %>% summarise(avg_price = mean(price))`` menghitung rata-rata harga per kategori.

7. `transmute()`

Mirip dengan ``mutate()``, tetapi hanya menyimpan kolom-kolom baru yang dihasilkan, membuang kolom asli yang tidak dimodifikasi.

Contoh: ``transmute(data, total = price * quantity)`` hanya menyimpan kolom ``total``.

Berguna untuk mengurangi ukuran data frame dengan fokus pada hasil transformasi.

2. **Data Kunjungan Perpustakaan**

NO	Bulan	Jumlah Pengunjung
1	January	90
2	February	86
3	Maret	97
4	April	83
5	Mei	85
6	Juni	85
7	Juli	65
8	Agustus	58
9	September	60
10	Oktober	93
11	Nopember	94
12	Desember	54

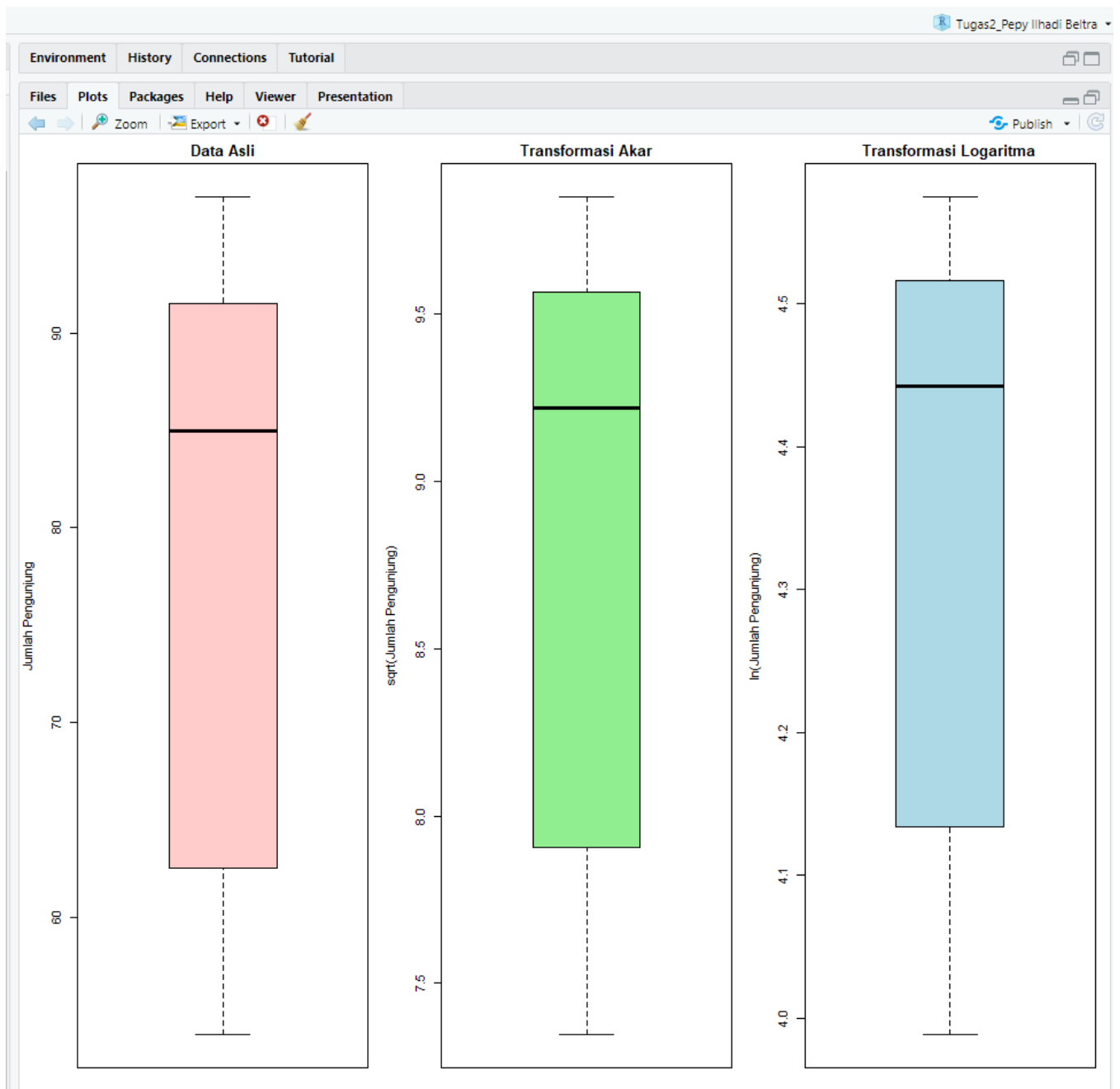
- a. Transformasi akar dan transformasi logaritma pada data tersebut menggunakan boxplot dan konsep transformasi tukey

- Source code pada R

```
Tugas_2_Visualisasi_Data_Peppy_Ilhadi_B.R x
Source on Save
Run
Source

1 # -----
2 # TRANSFORMASI DATA KUNJUNGAN PERPUSTAKAAN DAERAH
3 # -----
4
5 # 1. Input Data
6 data_kunjungan <- data.frame(
7   Bulan = c("Januari", "Februari", "Maret", "April", "Mei", "Juni",
8             "Juli", "Agustus", "September", "Oktober", "November", "Desember"),
9   Jumlah = c(90, 86, 97, 83, 85, 85, 65, 58, 60, 93, 94, 54)
10 )
11
12 # 2. Transformasi Tukey
13 # a. Transformasi akar (sqrt)
14 data_kunjungan$Akar <- sqrt(data_kunjungan$Jumlah)
15
16 # b. Transformasi logaritma natural (ln)
17 data_kunjungan$Logaritma <- log(data_kunjungan$Jumlah)
18
19 # 3. Visualisasi Boxplot
20 par(mfrow = c(1, 3), mar = c(4, 4, 2, 1)) # Atur layout plot
21
22 # Boxplot data asli
23 boxplot(data_kunjungan$Jumlah,
24         main = "Data Asli",
25         ylab = "Jumlah Pengunjung",
26         col = "#FFCCCB")
27
28 # Boxplot transformasi akar
29 boxplot(data_kunjungan$Akar,
30         main = "Transformasi Akar",
31         ylab = "sqrt(Jumlah Pengunjung)",
32         col = "#90EE90")
33
34 # Boxplot transformasi logaritma
35 boxplot(data_kunjungan$Logaritma,
36         main = "Transformasi Logaritma",
37         ylab = "ln(Jumlah Pengunjung)",
38         col = "#ADD8E6")
39
40 # Reset layout plot
41 par(mfrow = c(1, 1))
42
43 # 4. Tampilkan data hasil transformasi
44 print(data_kunjungan)
45
```

```
> print(data_kunjungan)
      Bulan Jumlah      Akar Logaritma
1 Januari      90 9.486833 4.499810
2 Februari     86 9.273618 4.454347
3 Maret       97 9.848858 4.574711
4 April       83 9.110434 4.418841
5 Mei         85 9.219544 4.442651
6 Juni        85 9.219544 4.442651
7 Juli        65 8.062258 4.174387
8 Agustus     58 7.615773 4.060443
9 September   60 7.745967 4.094345
10 Oktober    93 9.643651 4.532599
11 November   94 9.695360 4.543295
12 Desember   54 7.348469 3.988984
>
```



b) Langkah-langkah dalam melakukan Transformasi Data

1. Memasukkan Data

- Input data ke dalam data frame di R sesuai tabel.

2. Mengecek Distribusi Awal

- Gunakan boxplot untuk melihat sebaran data asli, outlier, dan skewness.

3. Melakukan Transformasi

- Terapkan transformasi akar (`sqrt()`) pada kolom jumlah pengunjung.

- Terapkan transformasi logaritma ($\log()$) pada kolom jumlah pengunjung.
 - Pastikan semua data bernilai positif sebelum menggunakan log.
4. Membandingkan Hasil
- Visualisasikan data hasil transformasi dengan boxplot.
 - Bandingkan bentuk sebaran, outlier, dan simetri data sebelum dan sesudah transformasi.
5. Interpretasi
- Pilih hasil transformasi yang membuat data lebih simetris dan variansi lebih stabil.
 - Transformasi yang tepat akan memudahkan analisis statistik lanjutan.

-----%%-----

Sumber Referensi:

BMP MSIM4310

<https://rpubs.com/ismailakbar/data-transformasi>