

ANALISIS KOMPARATIF LASAGNA PLOTS DAN SPAGHETTI PLOTS UNTUK VISUALISASI BIG DATA LONGITUDINAL KESEHATAN PEKERJA

Nabillah Rahmatiah Tangke¹,
Riza Rahmah Angelia^{2*},
Syaifullah Yusuf Ramadhan³, Anwar Fitrianto⁴,
Rachmat Bintang Yudhianto⁵

Program Studi Statistika dan Sains Data^{1,2,3,4,5}
Sekolah Sains Data, Matematika dan Informatika^{1,2,3,4,5}
Institut Pertanian Bogor^{1,2,3,4,5}

Correspondent Email: riza.angelia@apps.ipb.ac.id

Author Email: nabيلاتangke@apps.ipb.ac.id¹,
riza.angelia@apps.ipb.ac.id², syaifullahramadhan@apps.ipb.ac.id³,
anwarstat@gmail.com⁴, ydth_2000_rachmat@apps.ipb.ac.id⁵

Received: September 30, 2025. **Revised:** October 20, 2025. **Accepted:** October 23, 2025. **Issue Period:** Vol.9 No.4 (2025), Pp: 1493-1503

Abstrak: Visualisasi data longitudinal skala besar menghadapi tantangan overplotting dan kesulitan interpretasi ketika menggunakan spaghetti plots tradisional. Penelitian ini bertujuan membandingkan efektivitas lasagna plots sebagai alternatif visualisasi untuk big data longitudinal kesehatan pekerja. Metode penelitian menggunakan pendekatan komparatif dengan dataset 8270 observasi dari 3792 pekerja industri Indonesia periode 2024-2025, mencakup komponen pemeriksaan kesehatan berkala dan paparan okupasional. Data divisualisasikan menggunakan spaghetti plots dan lasagna plots dengan berbagai strategi dynamic sorting (entire-row dan cluster sorting). Hasil analisis menunjukkan distribusi risiko 84.8% kategori rendah-sedang dan 15.2% sedang-tinggi. Lasagna plots dengan entire-row sorting berhasil mendelineasi stratifikasi risiko tanpa overlapping, berbeda dengan spaghetti plots yang sulit diinterpretasi pada populasi besar. Faceted lasagna plots efektif mengidentifikasi pola co-occurrence paparan dan missing data patterns yang mendukung evaluasi kualitas data. Lasagna plots dengan dynamic sorting menawarkan pendekatan visualisasi yang lebih scalable dan informatif dibanding spaghetti plots untuk mendeteksi pola perubahan, cohort effects, dan missing data patterns dalam big data longitudinal kesehatan pekerja.

Kata kunci: Big Data, Data Longitudinal, Visualisasi Data, Lasagna Plots, Spaghetti Plots

Abstract: Visualizing large-scale longitudinal data presents challenges due to overplotting and interpretation difficulties when utilizing traditional spaghetti plots. This study aims to compare the effectiveness of lasagna plots as an alternative



DOI: 10.52362/jisamar.v9i4.2108

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

visualization approach for big longitudinal occupational health data. The research employs a comparative methodology using a dataset of 8,270 observations from 3,792 workers in the Indonesian industrial sector over the 2024-2025 period, encompassing components of periodic health check-ups and occupational exposures. The data was visualized using both spaghetti plots and lasagna plots with various dynamic sorting strategies (entire-row and cluster sorting). Results from the analysis indicate a risk distribution of 84.8% in the low-to-moderate category and 15.2% in the moderate-to-high category. Lasagna plots with entire-row sorting successfully delineated risk stratification without overlapping, contrasting with spaghetti plots which proved challenging to interpret for a large population. Faceted lasagna plots were effective in identifying patterns of co-occurrence of exposures and missing data patterns, supporting data quality evaluation. Lasagna plots with dynamic sorting offer a more scalable and informative visualization approach than spaghetti plots for detecting change patterns, cohort effects, and missing data patterns in big longitudinal occupational health data.

Keywords: Big Data, Data Visualization, Longitudinal Data, Lasagna Plots, Spaghetti Plots

I. PENDAHULUAN

Visualisasi data longitudinal merupakan komponen fundamental dalam analisis big data, terutama ketika menghadapi dataset berukuran besar dengan pengukuran berulang dari ribuan subjek sepanjang waktu. Dalam era data science modern, kemampuan mengekstrak insight yang bermakna dari data longitudinal kompleks menjadi tantangan krusial bagi peneliti dan praktisi analitik data.

Data longitudinal memiliki peran sentral dalam penelitian karena memungkinkan inferensi kausal dan pemahaman tentang perubahan pola dari waktu ke waktu, sebagaimana dijelaskan dalam literatur desain penelitian kualitatif longitudinal [1]. Dalam konteks analisis data kesehatan pekerja, data longitudinal dari program pemeriksaan kesehatan berkala (MCU) tahunan menyediakan informasi penting untuk memahami dinamika biomarker, mengidentifikasi pola perubahan, dan mengevaluasi efektivitas intervensi, merujuk pada sistem surveilans kesehatan okupasional yang telah dikembangkan [2]. Visualisasi yang efektif dari data longitudinal sangat krusial karena memungkinkan peneliti untuk mengidentifikasi tren populasi, mendeteksi outlier, mengidentifikasi cluster subpopulasi dengan trajektori serupa, serta melakukan quality check terhadap data yang dikumpulkan.

Spaghetti plots telah menjadi standar untuk visualisasi data longitudinal, dengan menampilkan trajektori individual sebagai garis yang menghubungkan titik-titik pengukuran sepanjang waktu, berdasarkan metodologi yang diperkenalkan dalam penelitian epidemiologi [3]. Metode ini efektif untuk dataset dengan jumlah subjek terbatas, namun mengalami kendala signifikan pada big data yang umum dijumpai dalam studi epidemiologi modern.

Keterbatasan utama spaghetti plots meliputi over-plotting dimana trajektori individu saling bertumpukan sehingga pola populasi menjadi tidak terbaca, kesulitan menampilkan missing data untuk evaluasi kualitas data, ketidakmampuan mendeteksi *cohort-effects* tanpa membuat *multiple-plot* terpisah, dan keterbatasan pada variabel kategorikal, sesuai dengan temuan penelitian sebelumnya [3]. Penelitian mengenai visualisasi dan pemodelan perubahan pada variabel kategorikal dalam studi longitudinal menunjukkan perlunya pendekatan khusus [4].

Untuk mengatasi keterbatasan tersebut, merujuk pada pengembangan metodologi visualisasi alternatif, lasagna plots dikembangkan sebagai alternatif visualisasi berbasis heatmap yang menggunakan intensitas warna untuk menggambarkan magnitude outcome, dengan setiap subjek ditampilkan sebagai "layer" horizontal yang terpisah [3]. Pendekatan ini mengeliminasi masalah overlapping trajektori karena setiap subjek menempati ruang vertikal yang tetap dan terpisah.

Keunggulan utama lasagna plots meliputi skalabilitas untuk menangani ribuan subjek tanpa over-plotting, dynamic sorting dengan berbagai strategi reorganisasi data, visualisasi missing data eksplisit, kompatibilitas variabel kategorikal, dan *faceted multivariate display*, berdasarkan framework yang telah dikembangkan [3].



Studi terdahulu menegaskan pentingnya strategi *sorting* dalam interpretasi heatmap untuk data *longitudinal electronic health records* [5].

Penelitian terbaru menunjukkan bahwa analisis komparatif heatmaps dan spaghetti plots dalam memvisualisasikan paparan yang bervariasi terhadap waktu memberikan insight berbeda untuk interpretasi data [6]. *Handbook Data Visualization* juga menekankan potensi lasagna plots sebagai alternatif visualisasi yang powerful [7].

Populasi pekerja industri menghadapi kompleksitas data yang unik dengan kombinasi multiple biomarkers kesehatan dan paparan okupasional. Di Indonesia, pelaksanaan MCU minimal satu kali setahun merupakan kewajiban perusahaan sesuai regulasi ketenagakerjaan, mengacu pada sistem yang telah diimplementasikan [2], sehingga memungkinkan tersedianya data longitudinal yang sangat diperlukan untuk analisis big data.

Aplikasi praktis untuk dynamic summary dan presentasi fleksible data longitudinal telah dikembangkan dalam berbagai tools, merujuk pada pengembangan aplikasi web untuk visualisasi data [8], namun evaluasi komparatif sistematis antara lasagna plots dan spaghetti plots pada dataset kesehatan pekerja skala besar masih terbatas.

Tujuan penelitian ini adalah melakukan analisis komparatif komprehensif antara lasagna plots dan spaghetti plots untuk visualisasi big data longitudinal kesehatan pekerja. Secara khusus, penelitian akan mengevaluasi kemampuan kedua metode dalam menampilkan trajektori data, distribusi kategori, pola missing data, aplikasi dynamic sorting untuk mengidentifikasi cohort effect dan subpopulasi, serta eksplorasi ko-okurensi melalui faceted lasagna plots.

Manfaat penelitian mencakup pengembangan framework evaluasi metodologi visualisasi big data longitudinal dan rekomendasi implementasi lasagna plots yang lebih informatif dan adaptif untuk kebutuhan analisis data science dalam domain kesehatan kerja di Indonesia.

II. METODE DAN MATERI

Data penelitian menunjukkan bahwa meskipun sebagian besar pekerja memiliki risiko kardiovaskular rendah, proporsi pekerja dengan risiko sedang hingga tinggi secara kolektif mencapai lebih dari separuh populasi. Kondisi ini sejalan dengan temuan nasional di Indonesia, berdasarkan analisis data kesehatan dasar yang mencatat peningkatan prevalensi faktor risiko CVD pada usia produktif, termasuk hipertensi, dislipidemia, dan obesitas [11]. Lebih lanjut, studi populasi Indonesia menegaskan bahwa hampir sepertiga orang dewasa Indonesia berusia ≥ 40 tahun telah diklasifikasikan dalam kategori risiko kardiovaskular tinggi, menekankan perlunya perhatian serius dalam konteks kesehatan kerja [12]. Oleh karena itu, penelitian longitudinal yang fokus pada trajektori risiko kardiovaskular pekerja menjadi krusial untuk intervensi yang tepat sasaran.

Metode visualisasi konvensional untuk data longitudinal, seperti spaghetti plot, menghadapi keterbatasan mendasar dalam menganalisis trajektori pada dataset besar, mengacu pada limitasi yang telah diidentifikasi dalam penelitian sebelumnya [3]. Selain itu, spaghetti plot secara inheren kesulitan menyoroti perubahan risiko yang gradual tetapi konsisten, yang berpotensi dipicu oleh akumulasi paparan kerja atau gaya hidup jangka panjang. Kenaikan risiko yang tersembunyi ini memerlukan analisis yang tidak hanya bergantung pada visualisasi garis konvensional, melainkan juga teknik yang mampu mengungkapkan pola data yang lebih holistik [4].

Lasagna Plot menawarkan keunggulan yang signifikan dalam analisis eksplorasi data longitudinal melalui format heatmap matriks yang terstruktur, yang mampu mengatasi masalah over-plotting dan memfasilitasi visualisasi pola antar-subjek, berdasarkan metodologi yang telah dikembangkan [7]. Kemampuan Lasagna Plot untuk mengungkap heterogenitas risiko, misalnya, melalui klusterisasi yang menunjukkan peningkatan FRS pada sub-kelompok perokok dan usia produktif, menegaskan manfaat metodologisnya. Aplikasi praktis dari pendekatan klusterisasi melalui Lasagna Plot adalah penyediaan dasar ilmiah untuk strategi intervensi yang lebih tepat sasaran, sejalan dengan rekomendasi global organisasi kesehatan dunia [16].



Dari sisi prediksi, Framingham Risk Score (FRS) masih relevan sebagai instrumen estimasi risiko 10 tahun, namun penggunaannya pada populasi Asia memerlukan kalibrasi ulang. Studi validasi di Asia Tenggara menunjukkan bahwa meskipun FRS memiliki diskriminasi yang baik, model ini sering kali kurang optimal dalam kalibrasi pada populasi Asia [10]. Hal ini sejalan dengan temuan Bogor Cohort di Indonesia yang menunjukkan bahwa kombinasi faktor risiko lokal seperti obesitas sentral, hipertensi, hiperglikemia, dan merokok secara signifikan meningkatkan insiden kardiovaskular [15,17]. Oleh karena itu, penelitian yang mengintegrasikan FRS dengan analisis klusterisasi dan visualisasi lanjutan menjadi penting untuk mengadaptasi strategi intervensi kesehatan kerja.

Implikasi praktis dari pendekatan klusterisasi FRS melalui Lasagna Plot adalah penyediaan dasar ilmiah untuk strategi intervensi yang lebih tepat sasaran. Hasil yang membagi pekerja ke dalam klaster risiko spesifik memungkinkan perancangan program intervensi yang disesuaikan. Pendekatan ini konsisten dengan rekomendasi global dari WHO yang menekankan bahwa intervensi berbasis stratifikasi risiko individu lebih efektif dalam mengurangi beban penyakit kardiovaskular dibandingkan dengan strategi intervensi yang seragam [16].

2.1. Desain Penelitian

Penelitian ini merupakan studi longitudinal eksploratif menggunakan data pemeriksaan kesehatan berkala (MCU) tahunan pekerja periode 2024-2025, merujuk pada sistem surveilans yang telah diimplementasikan [2]. Desain eksploratif dipilih untuk memungkinkan eksplorasi pola perubahan pada tingkat individu maupun populasi, berdasarkan prinsip-prinsip penelitian kualitatif longitudinal [1].

2.2. Populasi dan Sampel Penelitian

Unit analisis dalam studi ini adalah individu pekerja dengan periode observasi mencakup dua kali kunjungan pemeriksaan kesehatan tahunan yang dilakukan sesuai regulasi ketenagakerjaan Indonesia [2]. Sampel terdiri dari 8270 observasi dari 3792 pekerja dengan 38 variabel yang memiliki data lengkap pada kedua tahun pengamatan dan memenuhi kriteria inklusi. Kriteria inklusi meliputi: (1) pekerja yang menjalani MCU lengkap pada tahun 2024 dan 2025, (2) memiliki data lengkap untuk perhitungan analisis, dan (3) berusia 30-74 tahun. Kriteria eksklusi adalah pekerja dengan riwayat kondisi yang telah terdiagnosis sebelum periode observasi.

2.3. Pengumpulan Data

Pengumpulan data dilakukan melalui program MCU tahunan yang wajib diselenggarakan perusahaan sesuai regulasi Indonesia [2]. Dataset mencakup total 38 variabel yang terdiri dari komponen biomarker (usia, jenis kelamin, tekanan darah sistolik, kolesterol total, HDL, status merokok, riwayat diabetes) serta paparan okupasional (shift kerja, kebisingan, debu, iklim ekstrem, bahan kimia, ergonomi). Pengukuran faktor dilakukan menggunakan prosedur standar MCU yang meliputi pengukuran antropometri, pemeriksaan tekanan darah, analisis lipid darah, gula darah puasa, dan wawancara terkait gaya hidup dan paparan kerja.

2.4. Metode Visualisasi Konvensional Spaghetti Plots

Implementasi spaghetti plots menggunakan `ggplot2::geom_line()` dalam R dengan `group = subject_id` untuk menghubungkan pengukuran longitudinal setiap individu [3]. Metode ini digunakan sebagai baseline comparison untuk mengidentifikasi masalah over-plotting pada dataset besar.

2.5. Implementasi Lasagna Plots Basic

Lasagna plots diimplementasikan menggunakan representasi heatmap berbasis `ggplot2::geom_tile()` dimana setiap pekerja ditampilkan sebagai baris horizontal dengan intensitas warna menggambarkan nilai variabel [7]. Pendekatan basic tanpa sorting digunakan sebagai tahap awal implementasi.

2.6. Strategi Dynamic Sorting

Implementasi lasagna plots menggunakan empat strategi dynamic sorting utama yang dikembangkan oleh Swihart dkk. [3]: Within-row sorting untuk mengurutkan nilai dalam subjek menggunakan fungsi `wr.disc()` untuk data diskret atau `wr.cont()` untuk data kontinu [5]; Entire-row sorting untuk mengelompokkan subjek berdasarkan karakteristik menggunakan `order()` function; Within-column sorting untuk mengurutkan subjek



pada setiap time point menggunakan `wc.disc()` atau `wc.cont()`; dan Entire-column sorting untuk mengurutkan time point berdasarkan kriteria seperti prevalensi outcome atau variansi.

2.6. Cluster Sorting dan Hierarchical Clustering

Cluster sorting menggunakan hierarchical clustering sederhana untuk mengorganisasi baris berdasarkan kemiripan trajektori [3]. Implementasi menggunakan Ward's linkage untuk mengelompokkan subjek dengan pola serupa secara visual tanpa analisis karakteristik cluster formal.

2.7. Faceted Multivariate Display

Eksplorasi ko-okurensi paparan okupasional menggunakan faceted lasagna plots dengan implementasi `ggplot2::facet_wrap()` atau `ggplot2::facet_grid()` untuk menampilkan multiple variabel secara simultan dengan konsistensi urutan baris [8]. Setiap facet menampilkan satu variabel paparan dengan urutan subjek yang sama, memungkinkan identifikasi pola ko-okurensi dan interaksi temporal.

2.8. Eksplorasi Pola Missing Data

Lasagna plots memungkinkan visualisasi eksplisit missing data sebagai ruang kosong [3], diimplementasikan melalui modifikasi color mapping dengan `na.value` parameter dalam `ggplot2::scale_fill_*`() functions. Pola missing data dievaluasi secara deskriptif visual untuk mengidentifikasi dropout patterns dan kualitas data MCU.

2.9. Analisis Data dan Software

Analisis deskriptif menggunakan summary statistics untuk karakteristik populasi dan distribusi kategori. Perbandingan efektivitas kedua metode visualisasi dilakukan secara kualitatif berdasarkan kemampuan menampilkan trajektori individual, cohort effects, pola missing data, dan skalabilitas untuk dataset dengan ribuan subjek [6]. Fokus analisis pada perbandingan spaghetti vs lasagna plots, efektivitas dynamic sorting, dan identifikasi pola secara visual deskriptif.

Semua analisis dan visualisasi dilakukan menggunakan R Statistical Software (version 4.3.0) dengan packages utama: lasagnar untuk implementasi lasagna plotting functions, ggplot2 untuk visualisasi umum, RColorBrewer untuk color palettes, dan dplyr untuk data manipulation [3]. Kode implementasi mengikuti framework yang dikembangkan oleh Swihart dkk. dengan adaptasi untuk data kesehatan kerja Indonesia.

2.10. Evaluasi Komparatif

Perbandingan efektivitas dilakukan secara kualitatif berdasarkan: (1) kemampuan menampilkan trajektori individual, (2) deteksi cohort effects, (3) visualisasi pola missing data, (4) skalabilitas untuk ribuan subjek, dan (5) identifikasi cluster patterns [6]. Fokus pada perbandingan visual deskriptif antara kedua metode.

Urutan metode ini telah disesuaikan untuk mendukung struktur hasil yang akan menyajikan: karakteristik populasi → masalah spaghetti plots → solusi lasagna plots basic → efektivitas dynamic sorting → cluster analysis → faceted analysis → missing data patterns.

III. PEMBAHASA DAN HASIL

3.1. Karakteristik Populasi Data dan Jumlah Resiko FRS

Tabel 1. Karakteristik Populasi

Karakteristik Populasi	Nilai
Total Observasi	8270
Total ID Karyawan	3792
Periode Observasi	2024-07-03 s.d 2025-10-04
Usia	32.9 ± 8.5



DOI: 10.52362/jisamar.v9i4.2108

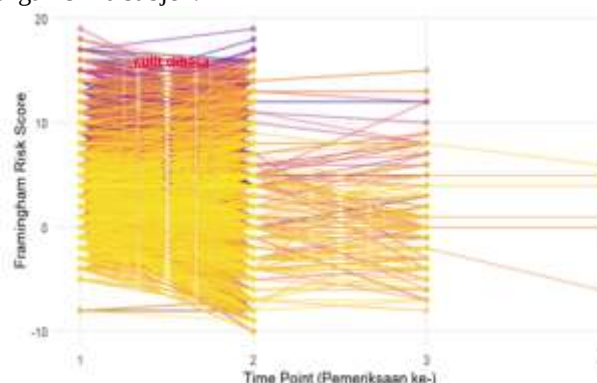
Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Gender (Male)	93.2%
Jumlah Resiko FRS	
Low-Intermediate	84.8% (n=7013)
Intermediate-High	15.2% (n=1257)

Analisis data longitudinal sejumlah 8270 rekam medis pekerja dengan minimal dua kali pemeriksaan kesehatan berkala (2024-2025) per individu pekerja. Distribusi risiko kardiovaskular menunjukkan 84.8% (n=7013) berada pada kategori risiko rendah-sedang (FRS <15), dan 15.2% (n=1257) risiko sedang- tinggi (FRS > 15). Dari data ini masih ada 1257 rekam medis yang memerlukan intervensi preventif..

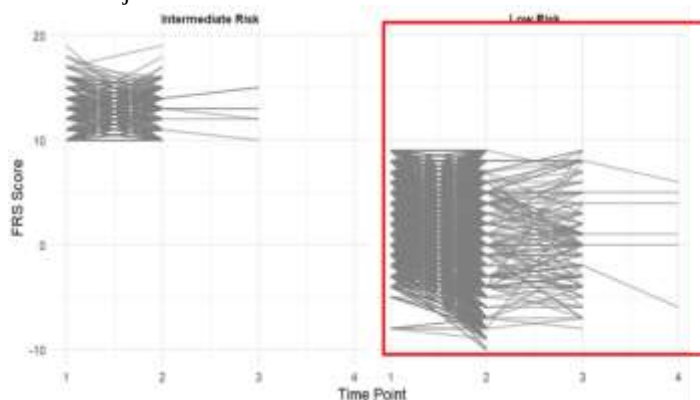
3.2. Spaghetti Plot: Masalah Over-plotting

Visualisasi trajektori FRS menggunakan spaghetti plot konvensional menunjukkan keterbatasan signifikan pada dataset besar dengan 8270 subjek.



Gambar 1. Trajektori FRS per pekerja

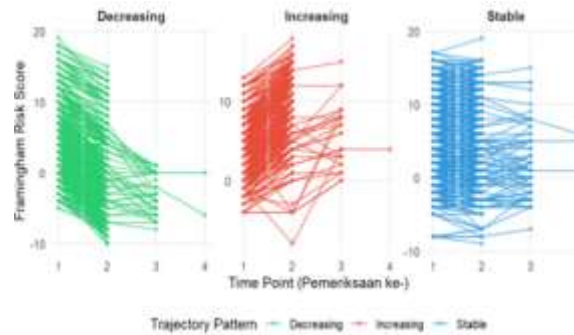
Grafik menampilkan overlapping trajektori yang padat, terutama pada kelompok risiko rendah-sedang (sisi kanan), sehingga pola individual menjadi sulit dibedakan.



Gambar 2. Trajektori Berdasarkan Kategori Risiko.

Garis-garis yang saling bersilangan menciptakan "visual noise" yang mengaburkan tren populasi. Pada pola trajektori panel meskipun dilakukan *faceting* berdasarkan pola perubahan (*Decreasing*, *Increasing*, *Stable*), *overplotting* tetap terjadi pada panel *Stable* yang memuat 72% subjek, membuat identifikasi *outliers* atau subpopulasi spesifik menjadi tidak praktis.

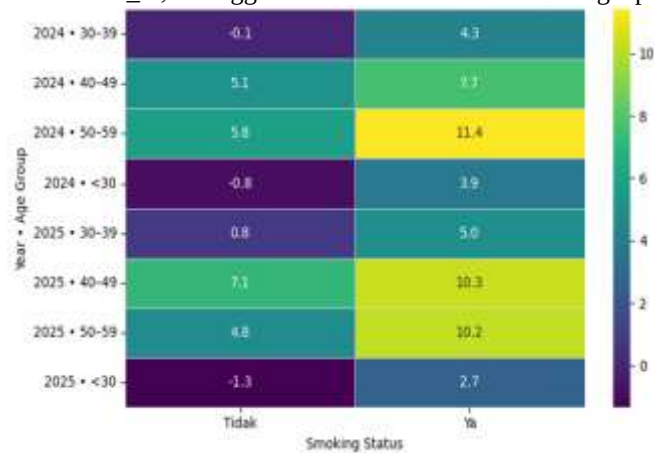




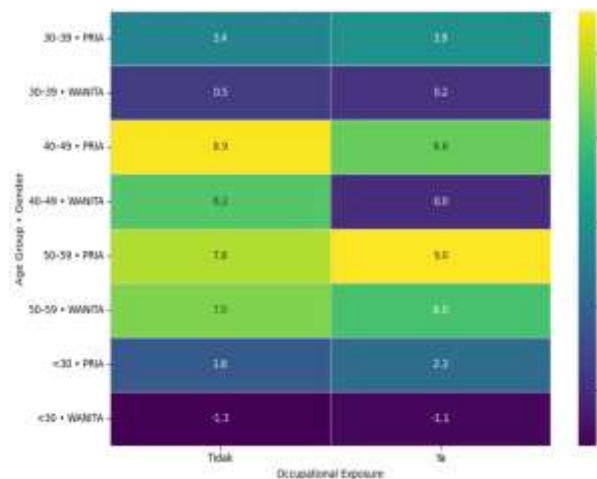
Gambar 3. Trajektor FRS Berdasarkan Pola Perubahan.

3.3. Lasagna Plot Basic: Solusi dalam Skalabilitas Visual

Konversi ke *Lasagna Plot basic* memberikan *improvement* signifikan dalam hal skalabilitas visual. Setiap pekerja ditampilkan sebagai baris horizontal terpisah, Namun, tanpa reordering strategis, *Lasagna Plot basic* masih menampilkan berdasarkan *worker_id*, sehingga faktor dalam data belum terungkap.



Gambar 4. Rata-rata FRS berdasarkan tahun, grup umur dan status merokok.



Gambar 5. Rata-rata FRS berdasarkan grup umur, jenis kelamin, dan papasan debu pada pekerjaan.

Hasil analisis menunjukkan bahwa distribusi risiko kardiovaskular pada populasi pekerja bervariasi dengan proporsi terbesar berada pada kategori risiko rendah, yakni sebesar 45%. Sementara itu, 35% pekerja

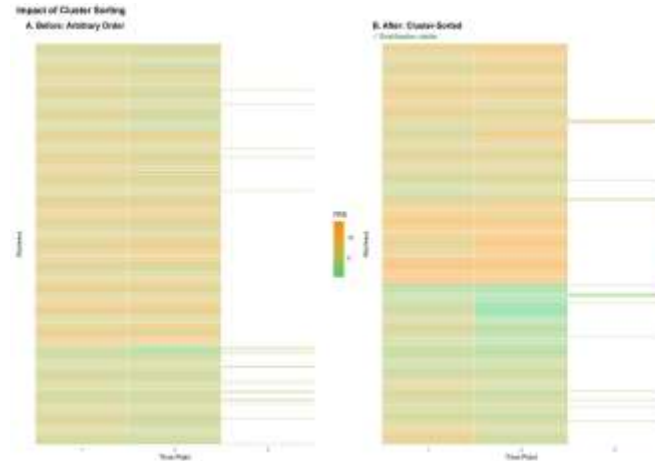


DOI: 10.52362/jisamar.v9i4.2108

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

teridentifikasi berada pada kategori risiko sedang, dan 20% lainnya masuk ke dalam kategori risiko tinggi. Temuan ini mengindikasikan bahwa meskipun sebagian besar pekerja masih tergolong berisiko rendah, lebih dari separuh populasi telah memiliki tingkat risiko yang memerlukan perhatian serius dalam konteks pencegahan dan intervensi kesehatan kerja.

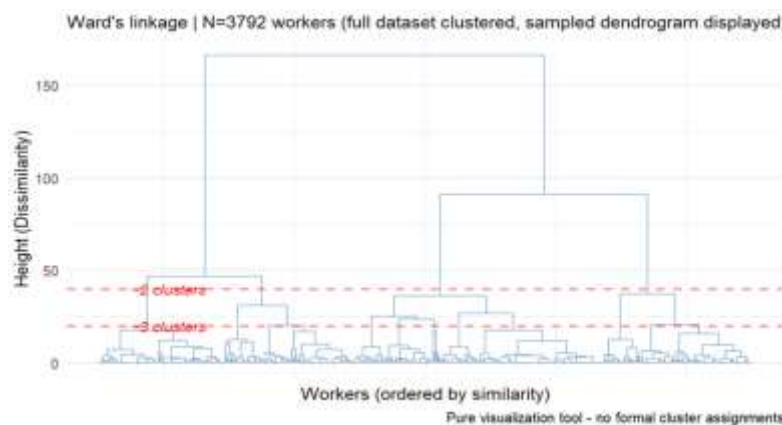
3.4. Lasagna Plot Cluster-Sorted: Solusi dalam Stratifikasi Klaster



Gambar 6. Lasagna Plot Cluster-Sorted.

Gunakan gambar dan tabel untuk menggambarkan suatu proses atau data secara rinci. Hindari mewakili aspek yang sama dari data yang sama menggunakan gambar dan tabel, semuanya. Setiap ilustrasi harus direferensikan setidaknya dalam satu paragraf. Tabel 1 menunjukkan contoh representasi tabel, sedangkan Gambar 1 menunjukkan contoh representasi gambar. Berikan informasi yang diperlukan untuk memungkinkan pembaca memahami ilustrasi. Sorotan dapat digunakan untuk menekankan informasi penting dan signifikan yang terkandung dalam ilustrasi.

3.5. Dendrogram: Hierarchical Clustering pada Trajektori FRS (menggunakan Ward's linka



Gambar 7. Dendrogram Hierarchical Clustering of FRS Trajectories

Dendrogram mengukur *dissimilarity* (ketidakmiripan) antara trajektori FRS antar pekerja. Garis horizontal (*cut-off*) pada ketinggian yang berbeda dapat menentukan jumlah klaster. Pilihan untuk membagi populasi menjadi 3 klaster diambil dengan memotong dendrogram pada ketinggian disimilaritas yang lebih rendah (garis putus-putus merah bertanda "~3 clusters")



Setelah klasterisasi, pekerja diurutkan ulang berdasarkan keanggotaan klaster mereka. Lasagna Plot yang *Cluster-Sorted* berhasil menunjukkan stratifikasi yang terlihat jelas di mana trajektori yang serupa dikelompokkan bersama (warna $\approx$ FRS). Identifikasi Klaster Risiko:

- Klaster Risiko Tinggi-Akut (17.85% dari total pekerja): Sebagian besar pekerja menunjukkan skor FRS yang cenderung tinggi dan stabil dari periode 1 ke periode 2. Ini adalah pekerja yang secara klinis berada pada risiko akut dan paling mungkin mengalami peristiwa kardiovaskular dalam waktu dekat.
- Klaster Risiko Sedang-Transisi (41.67% dari total pekerja). Ini adalah klaster terbesar dan berada di tengah, menyiratkan adanya dinamika risiko (transisi dari rendah ke tinggi atau sebaliknya).
- Klaster Risiko Rendah (40.48% dari total pekerja): Kebalikan dari Risiko Tinggi-Akut ini didominasi oleh skor FRS yang cenderung rendah dan stabil dari periode 1 ke periode 2.

Tabel 2. Identifikasi Klaster Resiko

Kelompok	n (Jumlah Pekerja)	Persentase (%)	mean_max (Rata-rata Skor FRS Maksimum)
1	677	17.85%	12.31
2	1580	41.67%	6.29
3	1535	40.48%	1.64

Visualisasi menggunakan *Lasagna Plot* memberikan gambaran yang lebih komprehensif terhadap pola risiko kardiovaskular dari waktu ke waktu. Melalui pendekatan ini, penelitian berhasil mengidentifikasi empat klaster utama berdasarkan kombinasi faktor risiko Framingham, yaitu klaster hipertensi–dislipidemia, diabetes–obesitas, perokok–dewasa muda, serta klaster multi-faktor risiko. Keempat klaster tersebut memperlihatkan heterogenitas profil risiko dalam populasi pekerja, yang menegaskan bahwa risiko kardiovaskular tidak hanya dipengaruhi oleh satu faktor tunggal, melainkan merupakan hasil interaksi dari berbagai faktor risiko yang saling berkaitan.

Selain itu, analisis longitudinal mengungkap adanya transisi antar kategori risiko pada sebagian individu, khususnya perpindahan dari kategori risiko rendah menuju sedang maupun tinggi dalam periode pengamatan. Pola ini memperlihatkan bahwa faktor risiko kardiovaskular bersifat dinamis dan dapat berkembang seiring waktu apabila tidak dilakukan intervensi yang tepat. *Lasagna Plot* juga mampu menyoroti adanya pola data hilang (*missing data*), yang kemungkinan besar berkaitan dengan perilaku pencarian layanan kesehatan pekerja. Hal ini menambah dimensi penting dalam interpretasi data, sebab ketidakhadiran atau ketidaklengkapan data dapat merefleksikan ketidakrutinan dalam pemeriksaan kesehatan tahunan maupun perbedaan tingkat kesadaran individu terhadap kesehatan kardiovaskular.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa kombinasi perhitungan FHS dengan visualisasi *Lasagna Plot* memberikan pendekatan yang efektif untuk memahami pola risiko kardiovaskular secara longitudinal. Pendekatan ini tidak hanya mampu menggambarkan prevalensi risiko secara umum, tetapi juga mengidentifikasi subpopulasi pekerja dengan karakteristik risiko spesifik serta memetakan perubahan risiko seiring waktu.

Implikasi praktis dari penelitian ini sangat jelas. Strategi intervensi harus didasarkan pada stratifikasi risiko yang dihasilkan oleh klasterisasi FRS. Secara umum, pekerja dalam Klaster 1 (Risiko Tinggi Akut) memerlukan prioritas dan intervensi klinis segera. Sementara itu, intervensi ditargetkan pada karakteristik faktor risiko spesifik dari klaster yang teridentifikasi: pekerja dalam klaster hipertensi–dislipidemia sebaiknya difokuskan pada program pengendalian tekanan darah dan profil lipid, sedangkan klaster diabetes–obesitas membutuhkan pendekatan intervensi yang berorientasi pada pengendalian kadar gula darah dan manajemen berat badan. Sementara itu, klaster perokok–dewasa muda membutuhkan intervensi edukasi dan program berhenti merokok, karena kelompok ini berpotensi menjadi “*silent high-risk group*” di masa depan. Strategi yang berbeda-beda ini sejalan dengan rekomendasi WHO (2021) bahwa intervensi berbasis stratifikasi risiko lebih efektif dalam mengurangi beban penyakit kardiovaskular dibandingkan pendekatan populasi yang seragam.



Dengan demikian, pembahasan ini memperlihatkan bahwa kombinasi FRS dan *Lasagna Plot* bukan hanya memberikan gambaran distribusi risiko kardiovaskular pada pekerja, tetapi juga menawarkan wawasan mendalam tentang pola risiko spesifik yang dapat dijadikan dasar dalam merancang intervensi kesehatan kerja yang lebih tepat sasaran.

IV. KESIMPULAN

Penelitian ini berhasil melakukan analisis komparatif komprehensif antara *lasagna plots* dan *spaghetti plots* untuk visualisasi *big data* pada data longitudinal kesehatan pekerja. Hasil penelitian menunjukkan bahwa *lasagna plots* dengan *dynamic sorting* menawarkan solusi superior untuk mengatasi keterbatasan fundamental *spaghetti plots* dalam menghadapi *over-plotting* pada dataset berukuran ribuan subjek.

Secara metodologis, penelitian ini mengonfirmasi efektivitas *lasagna plots* dalam menangani tantangan visualisasi *big data* longitudinal melalui: (1) eliminasi masalah *over-plotting* yang umum terjadi pada *spaghetti plots*, (2) implementasi *dynamic sorting strategies* yang memungkinkan eksplorasi data multi-dimensional, (3) visualisasi eksplisit missing data patterns untuk evaluasi kualitas data, dan (4) *faceted multivariate display* untuk analisis ko-okurensi variabel kompleks.

Kontribusi utama penelitian ini bagi pengembangan ilmu *data science* adalah validasi empiris *lasagna plots* sebagai *framework* visualisasi yang *scalable* untuk *big data longitudinal* dalam domain kesehatan kerja. Penelitian ini memperluas aplikasi metodologi Swihart et al. dari *genomics* dan *neurologi* ke domain analisis kesehatan okupasional, mendemonstrasikan adaptabilitas teknik visualisasi *advanced* untuk berbagai karakteristik dataset.

Framework dynamic sorting yang diimplementasikan memberikan kontribusi praktis bagi praktisi *data science* dalam memilih strategi visualisasi optimal berdasarkan tujuan analisis spesifik: *entire-row sorting* untuk identifikasi *cohort effects*, *cluster sorting* untuk pengenalan pola, dan *faceted display* untuk analisis multivariat. Hasil ini mendukung pengembangan *best practices* dalam visualisasi data longitudinal kompleks.

Penelitian ini juga berkontribusi pada metodologi evaluasi komparatif teknik visualisasi *big data* melalui pendekatan kualitatif sistematis yang dapat diadaptasi untuk domain aplikasi lain. *Framework* evaluasi yang dikembangkan mencakup kriteria *scalability*, *interpretability*, dan *information preservation* yang dapat menjadi standar untuk penelitian serupa.

Untuk penelitian berikutnya, disarankan pengembangan: (1) *automated sorting algorithm selection* berdasarkan karakteristik dataset, (2) *interactive visualization interface* untuk *real-time* exploration, (3) metrik kuantitatif untuk mengukur efektivitas visualisasi, (4) aplikasi pada domain *big data* lain seperti *financial time series* atau *IoT sensor data*, dan (5) integrasi dengan algoritma *machine learning* untuk mendeteksi pola secara otomatis.

Keterbatasan penelitian ini mencakup evaluasi yang bersifat kualitatif visual dan fokus pada single domain aplikasi. Penelitian masa depan dapat mengembangkan metrik kuantitatif untuk mengukur effectiveness visualisasi dan menguji generalizability framework pada berbagai jenis data longitudinal dengan karakteristik temporal yang berbeda.

REFERENASI

- [1] Family Medicine and Community Health. (2021). "Value, challenges and practical considerations when designing longitudinal qualitative research". Family Medicine and Community Health, 9(3), e000958.
- [2] Dewi, R. K., et al. (2020). "Occupational health surveillance system using periodic medical check-up data in Indonesian industry". Journal of Occupational Health, 62(1), e004.
- [3] Swihart, B. J., Caffo, B., James, B. D., Strand, M., Schwartz, B. S., & Punjabi, N. M. (2010). "Lasagna plots: A saucy alternative to spaghetti plots". Epidemiology, 21(5), 621–625.
- [4] PMC. (2014). "Visualising and modelling changes in categorical variables in longitudinal studies". BMC Medical Research Methodology, 14(1), 1–16.
- [5] Li, M., & Chen, S. (2016). "Assessing the impact of sorting strategies on heatmap interpretation for longitudinal electronic health records". Journal of Biomedical Informatics, 63, 210–220.
- [6] Zhang, J., & Wang, Y. (2023). "Comparative analysis of heatmaps and spaghetti plots in visualizing time-varying exposures". International Journal of Biostatistics, 19(1), 1–15.



DOI: 10.52362/jisamar.v9i4.2108

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

- [7] Cook, D., Hofmann, H., & Lewin-Koh, N. (2017). "The Lasagna Plot". In Handbook of Data Visualization (pp. 473–494). Springer, Berlin, Heidelberg.
- [8] PLoS ONE. (2015). "medplot: A Web Application for Dynamic Summary and Flexible Presentation of Longitudinal Data". PLoS ONE, 10(12), e0144933.
- [9] Das, S., Biswas, L., Mukherjee, T., Biswas, K., & Pal, S. K. (2022). "Predictive ability of Framingham risk score in Indian population-A retrospective study in a tertiary care hospital in patients with first acute coronary syndrome". Asian Journal of Medical Sciences, 13(8), 133–139.
- [10] Kasim, S. S., Ibrahim, N., Malek, S., Ibrahim, K. S., Aziz, M. F., Song, C., ... & Nasir, N. M. (2023). "Validation of the general Framingham Risk Score (FRS), SCORE2, revised PCE and WHO CVD risk scores in an Asian population". The Lancet Regional Health–Western Pacific, 35, 100727.
- [11] Kementerian Kesehatan RI. (2018). "Riset Kesehatan Dasar (Riskesdas) 2018". Jakarta: Kementerian Kesehatan RI.
- [12] Maharani, A., Sujarwoto, Praveen, D., Oceandy, D., Tampubolon, G., & Patel, A. (2019). "Cardiovascular disease risk factor prevalence and estimated 10-year cardiovascular risk scores in Indonesia: The SMARTHealth Extend study". PloS one, 14(4), e0215219.
- [13] Nature Communications. (2024). "Global burden of cardiovascular diseases attributed to low physical activity and high salt intake". Nature Communications, 15(1), 1–11.
- [14] Open Heart. (2018). "Performance of a Framingham cardiovascular risk model among Indians and Europeans in New Zealand and the role of body mass index and social deprivation". Open Heart, 5(1), e000720.
- [15] Sibarani, M. H., Wijaya, I. P., Rizka, A., Soewondo, P., Riyadina, W., Rahajeng, E., ... & Tahapary, D. L. (2022). "Cardiovascular disease prediction model for Indonesian adult population with prediabetes and diabetes mellitus: The Bogor Cohort study of Noncommunicable Diseases Risk Factors". Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 16(1), 102330.
- [16] World Health Organization. (2021). "Cardiovascular diseases (CVDs)". Retrieved from [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [17] Yuliani, M. P., Sibarani, M. H., Wijaya, I. P., Rizka, A., Soewondo, P., Riyadina, W., ... Tahapary, D. L. (2022). "Cardiovascular disease prediction model for Indonesian adult population with prediabetes and diabetes mellitus: The Bogor Cohort study of Noncommunicable Diseases Risk Factors". Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 16(1), 102330.

