

**ANALISIS KORESPONDENSI SEDERHANA
UNTUK MENGETAHUI HUBUNGAN ANTARA PERBEDAAN
GENERASI, USIA DAN NEGARA TERHADAP TINGKAT BUNUH DIRI**



**Dosen Pengampu:
Prof. Dr. Dra. Titin Siswantining, D.E.A.**

**Disusun oleh:
Shinta Chandra (2206053940)**

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS INDONESIA
2024**

Bagian 1. Pendahuluan

Tingkat bunuh diri merupakan isu kesehatan global yang kompleks dan beragam, yang dipengaruhi oleh berbagai faktor sosial, ekonomi, dan budaya. Generasi, usia, dan negara adalah beberapa variabel penting yang dapat memengaruhi pola dan tren tingkat bunuh diri di suatu populasi. Generasi yang berbeda mungkin memiliki pengalaman dan tuntutan sosial yang berbeda, sedangkan usia dapat memengaruhi faktor risiko dan protektif individu. Selain itu, faktor-faktor sosial, ekonomi, dan budaya yang unik untuk setiap negara juga dapat memengaruhi prevalensi dan pola bunuh diri. Oleh karena itu, memahami hubungan antara perbedaan generasi, usia, dan negara dengan tingkat bunuh diri sangat penting untuk merancang intervensi yang efektif dalam mencegah kasus bunuh diri.

Analisis korespondensi sederhana merupakan salah satu metode statistik yang digunakan untuk mengukur hubungan antara dua variabel kategorik yang disajikan dalam bentuk tabel kontingensi. Dalam konteks ini, teknik analisis korespondensi sederhana menjadi alat yang berguna untuk mengeksplorasi dan menganalisis hubungan antara perbedaan generasi, usia, dan negara terhadap tingkat bunuh diri.

Dalam proyek akhir ini, data yang digunakan bersumber dari **Kaggle: Suicide Rates Overview 1985 to 2016**. Data ini berisi angka bunuh diri dari berbagai negara yang dikumpulkan dari tahun 1985 sampai 2016. Berikut cuplikan data mentahnya:

country	year	sex	age	suicides_no	population	suicides/100,000
Albania	1987	male	15-24 years	21	312900	6.71
Albania	1987	male	35-54 years	16	308000	5.19
Albania	1987	female	15-24 years	14	289700	4.83
Albania	1987	male	75+ years	1	21800	4.59
Albania	1987	male	25-34 years	9	274300	3.28
Albania	1987	female	75+ years	1	35600	2.81

Data tersebut memiliki 27.820 observasi dan 12 kolom. Namun penulis akan mengambil kolom **'age'**, **'gender'**, **'country'**, **'suicides_no'** dan **'generation'** untuk keperluan analisis data. Keterangan variabel/kolom sebagai berikut:

1. Age: kelompok umur yang terdiri dari '25-34 years', '55-74 years', '35-54 years', '5-14 years', '75+ years' dan '15-24 years'.
2. Gender: jenis kelamin terdiri dari 'female' dan 'male'.
3. Country: negara 'United States', 'Kiribati', 'Republic of Korea', 'Guatemala', 'United Arab Emirates', 'Albania'.
4. Suicides_no: angka bunuh diri (satuan individu).
5. Generation: berisi generasi 'Boomers', 'Generation X', 'Silent', 'G.I. Generation', 'Millenials' dan 'Generation Z'.

G.I. Generation (1901-1927)

Generasi ini juga dikenal sebagai "Greatest Generation" karena mereka tumbuh besar selama masa Depresi Besar dan berjuang atau bekerja dalam Perang Dunia II. Mereka dikenal karena ketahanan dan rasa tanggung jawab kolektif.

Silent Generation (1928-1945)

Generasi ini tumbuh selama masa Depresi Besar dan Perang Dunia II. Mereka dikenal sebagai "Silent Generation" karena mereka cenderung konservatif dan tidak banyak bicara tentang masalah sosial.

Baby Boomers (1946-1964)

Generasi ini lahir setelah Perang Dunia II selama periode peningkatan kelahiran yang disebut "baby boom". Mereka mengalami kemakmuran ekonomi setelah perang dan menjadi pendorong utama dalam perubahan sosial selama 1960-an dan 1970-an.

Generation X (1965-1980)

Generasi ini tumbuh selama periode perubahan ekonomi yang cepat, peningkatan perceraian, dan teknologi baru. Mereka sering dianggap sebagai generasi yang mandiri dan skeptis.

Millennials (1981-1996)

Juga dikenal sebagai Generasi Y, mereka tumbuh bersama teknologi digital, internet, dan media sosial. Mereka sering dikaitkan dengan nilai-nilai inklusivitas dan keberagaman.

Generation Z (1997-2012)

Generasi ini lahir dalam dunia yang sepenuhnya digital. Mereka sering disebut sebagai "Digital Natives" karena tidak pernah mengenal dunia tanpa internet.

Semua variabel yang diambil adalah kategorik kecuali suicides_no (float/ desimal). Variabel suicides_no akan menjadi data kita dalam menganalisis variabel kategorik generasi, usia dan negara dengan korespondensi sederhana. Untuk lebih lengkapnya akan dijelaskan pada tahapan selanjutnya.

Bagian 2. Pre-Processing Dan Analisis Deskriptif Untuk Visualisasi Data

Tahapan ini dilakukan dengan bantuan program google colab.

Pre-processing

Berikut langkah-langkah yang dilakukan dalam pre-processing dataset:

1. Import dataset dan package yang diperlukan

```
1 #import package yang dibutuhkan
2 import pandas as pd
3 import seaborn as sns
4 import matplotlib.pyplot as plt

[ ] 1 #Import dataset
2 df = pd.read_csv('/content/master.csv')
3 df
```

	country	year	sex	age	suicides_no	population	suicides/100k pop	country-year	HDI for year	gdp_for_year (\$)	gdp_per_capita (\$)	generation
0	Albania	1987	male	15-24 years	21	312900	6.71	Albania1987	NaN	2,156,624,900	796	Generation X
1	Albania	1987	male	35-54 years	16	308000	5.19	Albania1987	NaN	2,156,624,900	796	Silent
2	Albania	1987	female	15-24 years	14	289700	4.83	Albania1987	NaN	2,156,624,900	796	Generation X
3	Albania	1987	male	75+ years	1	21800	4.59	Albania1987	NaN	2,156,624,900	796	G.I. Generation
4	Albania	1987	male	25-34 years	9	274300	3.28	Albania1987	NaN	2,156,624,900	796	Boomers
...

2. Mengambil kolom yang diperlukan

```

1 #mengambil kolom yang diinginkan
2 kolom = ["country", "sex", "age", "generation", "suicides_no"]

[ ] 1 selected_df = df[kolom]
2 selected_df

```

	country	sex	age	generation	suicides_no
0	Albania	male	15-24 years	Generation X	21
1	Albania	male	35-54 years	Silent	16
2	Albania	female	15-24 years	Generation X	14
3	Albania	male	75+ years	G.I. Generation	1
4	Albania	male	25-34 years	Boomers	9
...
27815	Uzbekistan	female	35-54 years	Generation X	107
27816	Uzbekistan	female	75+ years	Silent	9
27817	Uzbekistan	male	5-14 years	Generation Z	60
27818	Uzbekistan	female	5-14 years	Generation Z	44
27819	Uzbekistan	female	55-74 years	Boomers	21

3. Memilih 6 negara berdasarkan pertimbangan

```

[ ] 1 #memilih negara 6 negara yang akan diambil
2 selected_countries = ['United States', 'Kiribati', 'Republic of Korea', 'Guatemala', 'United Arab Emirates']
3 subset_df = selected_df[selected_df['country'].isin(selected_countries)]
4

```

	country	sex	age	generation	suicides_no
0	Albania	male	15-24 years	Generation X	21
1	Albania	male	35-54 years	Silent	16
2	Albania	female	15-24 years	Generation X	14
3	Albania	male	75+ years	G.I. Generation	1
4	Albania	male	25-34 years	Boomers	9
...
27215	United States	female	25-34 years	Millenials	1444
27216	United States	female	15-24 years	Millenials	1132
27217	United States	female	75+ years	Silent	540
27218	United States	male	5-14 years	Generation Z	255
27219	United States	female	5-14 years	Generation Z	158

1572 rows x 5 columns

4. Mengambil 500 observasi secara acak untuk mempermudah analisis data.

```

1 # Mengambil 500 observasi secara acak
2 df_500 = subset_df.sample(n=500, random_state=42)
3 df_500

```

	country	sex	age	generation	suicides_no
20594	Republic of Korea	male	25-34 years	Millenials	988
20284	Republic of Korea	female	55-74 years	G.I. Generation	136
26446	United Arab Emirates	female	35-54 years	Boomers	4
10974	Guatemala	female	35-54 years	Boomers	25
10954	Guatemala	female	5-14 years	Millenials	5
...
10752	Guatemala	male	75+ years	G.I. Generation	4
20510	Republic of Korea	female	25-34 years	Generation X	943
20496	Republic of Korea	female	75+ years	Silent	743
20554	Republic of Korea	male	75+ years	Silent	1137
109	Albania	male	15-24 years	Generation X	27

500 rows x 5 columns

5. Mengurutkan observasi dari kelompok umur terkecil hingga terbesar.

```

1 #mengurutkan data dari kelompok umur terkecil hingga terbesar
2 kategori_urutan = [
3     '5-14 years',
4     '15-24 years',
5     '25-34 years',
6     '35-54 years',
7     '55-74 years',
8     '75+ years'
9 ]
10 df_500['age'] = pd.Categorical(df_500['age'], categories=kategori_urutan, ordered=True)
11
12 # Mengurutkan DataFrame berdasarkan kolom 'age'
13 df_500 = df_500.sort_values(by='age')
14
15 # Menampilkan DataFrame setelah diurutkan
16 print(df_500)

```

	country	sex	age	generation	suicides_no
27207	United States	female	5-14 years	Generation Z	151
10727	Guatemala	female	5-14 years	Generation X	2
11050	Guatemala	male	5-14 years	Generation Z	6
14088	Kiribati	female	5-14 years	Millenials	1
10798	Guatemala	female	5-14 years	Millenials	4
...
20458	Republic of Korea	male	75+ years	Silent	722
11025	Guatemala	female	75+ years	Silent	1
20329	Republic of Korea	female	75+ years	G.I. Generation	91
27157	United States	female	75+ years	Silent	403
20554	Republic of Korea	male	75+ years	Silent	1137

[500 rows x 5 columns]

6. Reset index untuk mengatur ulang dari 1

```

1 #mereset index dataframe df_500
2 df_500.reset_index(drop=True, inplace=True)

```

7. Statistika deskriptif

suicides_no	
count	500.000000
mean	785.670000
std	1620.218503
min	0.000000
25%	5.000000
50%	38.000000
75%	744.250000
max	11455.000000

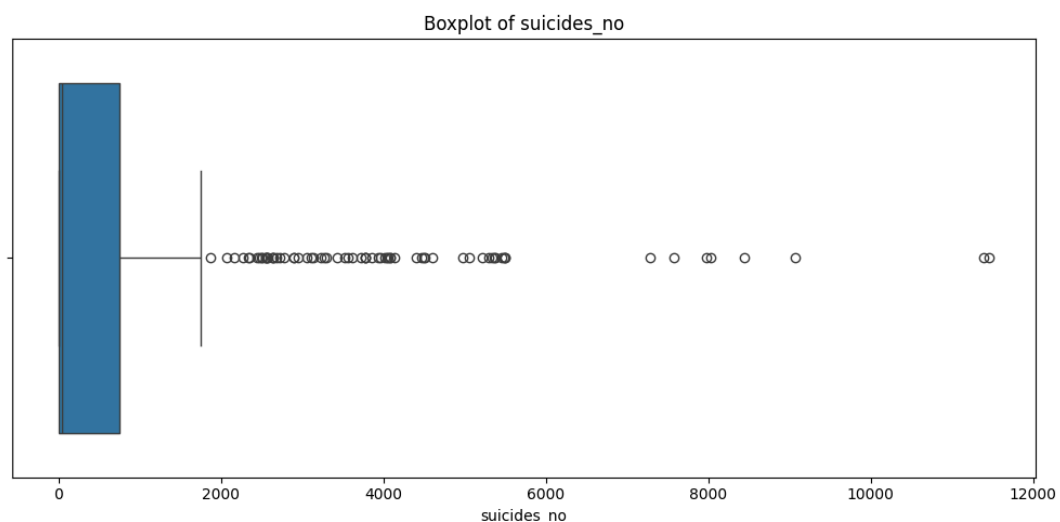
8. Mengecek missing value

[]	1 #mengecek missing value
	2 df_500.isnull().sum()
⇒	country 0
	sex 0
	age 0
	generation 0
	suicides_no 0
	dtype: int64

Visualisasi Data

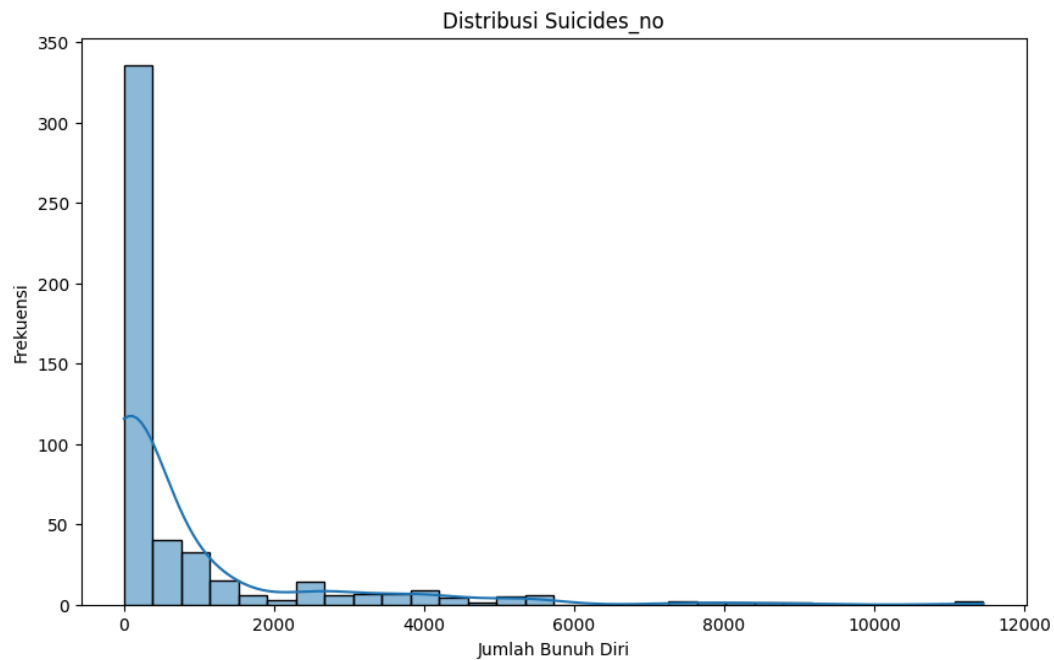
Tahapan ini dilakukan untuk memperoleh *insight* berharga dari data untuk masing-masing variabel yang digunakan.

1. Membuat box plot untuk kolom suicides_no



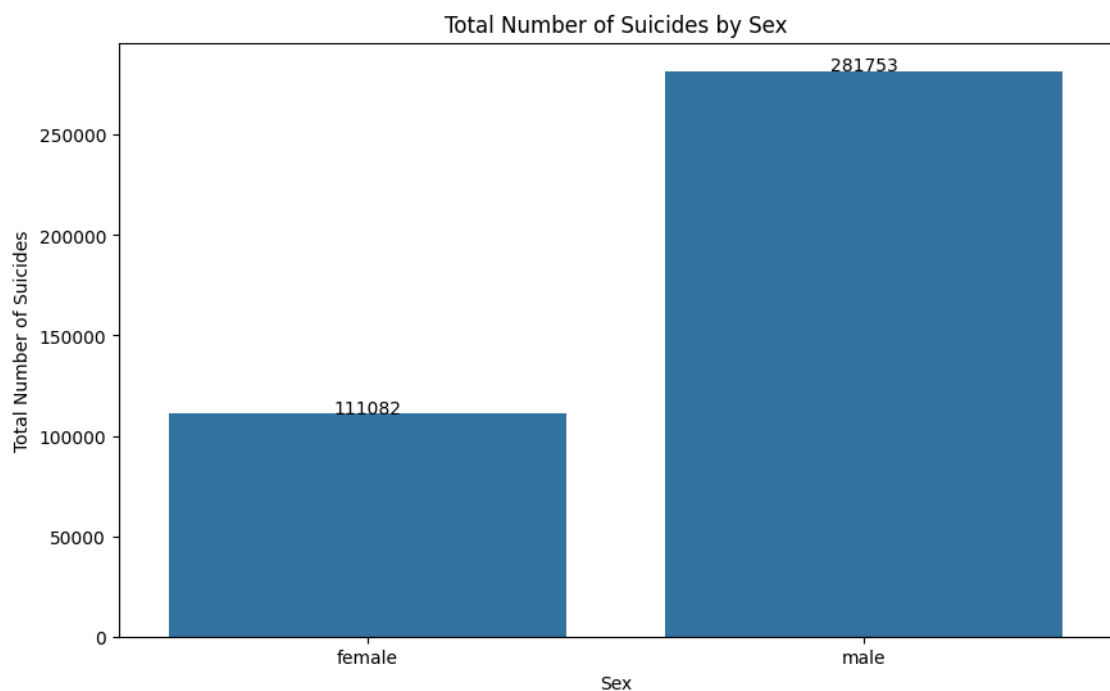
Terlihat banyak sekali outlier pada kolom suicides_no. Hal ini memberikan beberapa informasi ada negara yang memiliki angka kasus bunuh jauh lebih tinggi daripada lainnya. Penulis memutuskan untuk tidak menghapus outlier karena hal itu dapat memberikan informasi penting untuk analisis lebih lanjut.

2. Histogram untuk distribusi angka bunuh diri



Distribusi dari angka bunuh diri menunjukkan *skewness* ke arah kanan (positif skewed), yang berarti sebagian besar data berkelompok di sebelah kiri (dekat nol), dan hanya sedikit data yang tersebar ke arah kanan dengan nilai yang lebih tinggi. Hal ini berarti mayoritas observasi memiliki jumlah bunuh diri yang relatif rendah. Ada beberapa negara atau kategori yang memiliki jumlah bunuh diri sangat tinggi, tetapi ini merupakan minoritas dalam dataset.

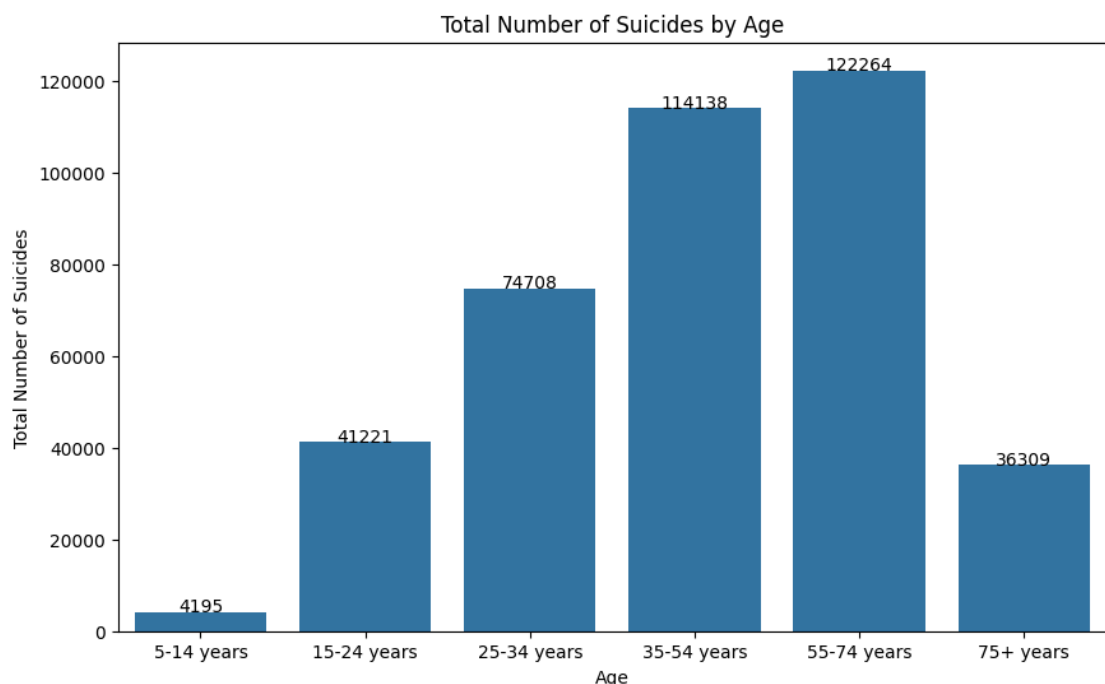
3. Bar plot banyaknya angka bunuh diri berdasarkan gender



Dilakukan untuk melihat banyaknya angka bunuh diri berdasarkan perbedaan gender. Dari plot di atas terlihat bahwa angka bunuh diri pada pria jauh lebih tinggi dibandingkan dengan wanita. Pria memiliki lebih dari dua kali lipat angka bunuh diri dibandingkan wanita dalam dataset ini. Jumlah kasus untuk perempuan = 111082 sementara laki-laki = 28175.

Pria sering kali mengalami tekanan sosial dan ekonomi yang tinggi, seperti menjadi pencari nafkah utama, yang dapat meningkatkan risiko stres dan depresi. Pria mungkin cenderung kurang mencari bantuan untuk masalah kesehatan mental karena stigma sosial yang terkait dengan menunjukkan kelemahan atau kerentanan.

4. Bar plot banyaknya angka bunuh diri berdasarkan age.



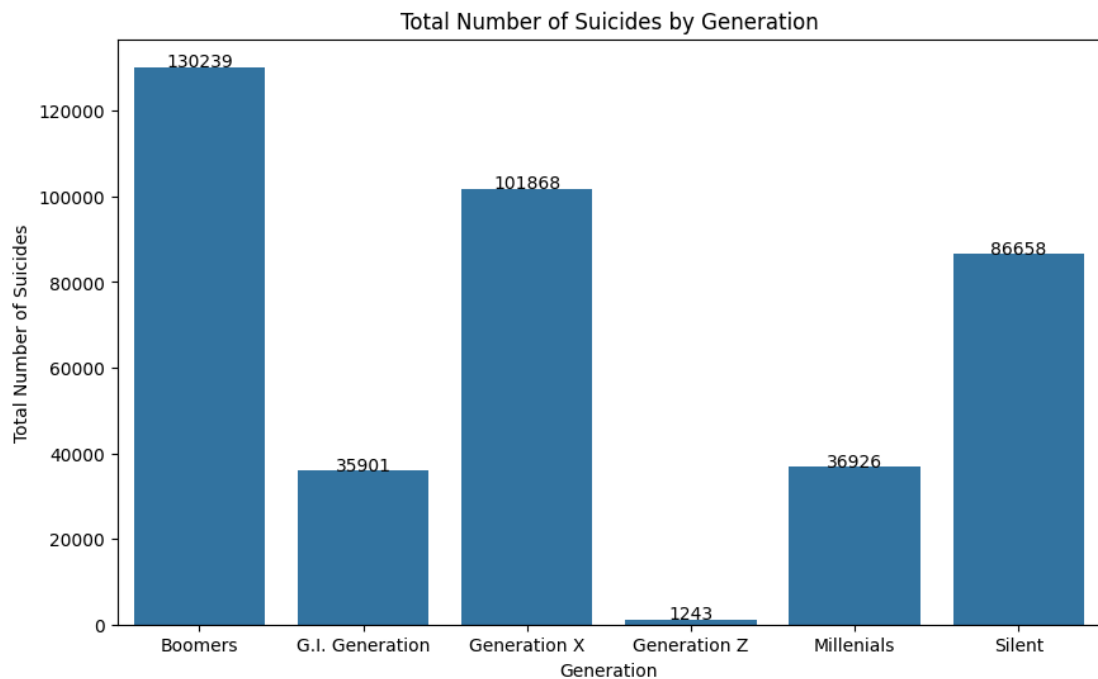
Dilakukan untuk melihat banyaknya angka bunuh diri berdasarkan perbedaan kelompok umur. Berikut interpretasinya:

- 5-14 years: Terdapat 4,195 kasus bunuh diri.
- 15-24 years: Terdapat 41,221 kasus bunuh diri.
- 25-34 years: Terdapat 74,708 kasus bunuh diri.
- 35-54 years: Terdapat 114,138 kasus bunuh diri.
- 55-74 years: Terdapat 122,264 kasus bunuh diri.
- 75+ years: Terdapat 36,309 kasus bunuh diri.

Angka bunuh diri meningkat seiring bertambahnya usia, mencapai puncak pada kelompok umur 55-74 tahun. Setelah itu, angka bunuh diri menurun pada kelompok umur 75+ tahun. Kelompok usia 55-74 tahun memiliki angka bunuh diri tertinggi, diikuti oleh kelompok 35-54 tahun.

Hal ini mungkin disebabkan oleh meningkatnya tekanan hidup seiring bertambahnya usia. Beban kerja, tanggung jawab keluarga, dan kurangnya pencapaian hidup dapat menjadi faktor penyebab.

5. Bar plot banyaknya angka bunuh diri berdasarkan Generation.

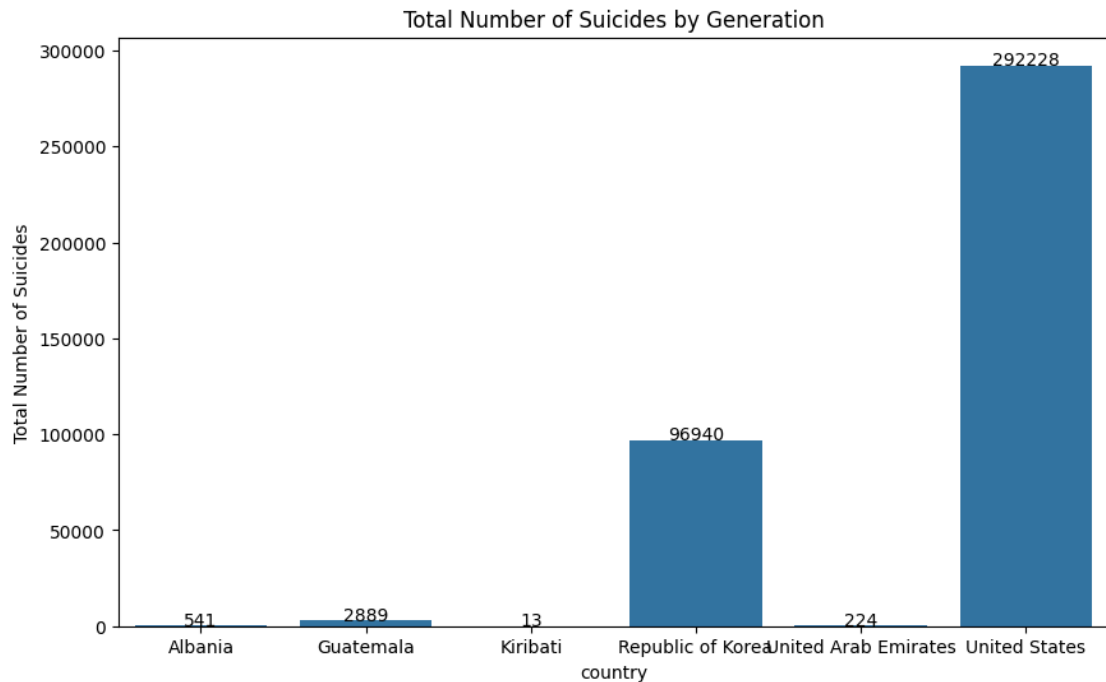


Dilakukan untuk melihat banyaknya angka bunuh diri berdasarkan perbedaan generasi. Berikut interpretasinya:

- Boomers: Terdapat 130,239 kasus bunuh diri.
- G.I. Generation: Terdapat 35,901 kasus bunuh diri.
- Generation X: Terdapat 101,868 kasus bunuh diri.
- Generation Z: Terdapat 1,243 kasus bunuh diri.
- Millennials: Terdapat 36,926 kasus bunuh diri.
- Silent: Terdapat 86,658 kasus bunuh diri.

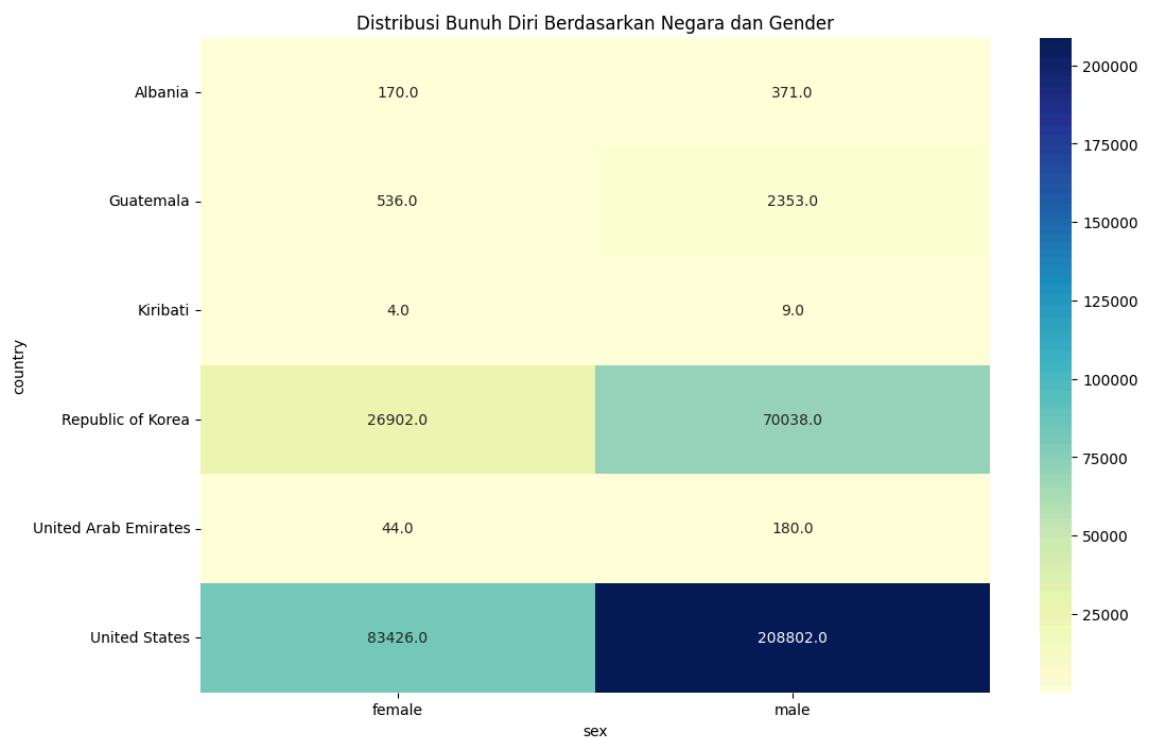
Generasi Baby Boomers memiliki angka bunuh diri tertinggi. Generasi X berada di posisi kedua dalam hal angka bunuh diri. Generasi G.I. dan Generasi Milenial memiliki angka bunuh diri yang hampir sama, tetapi jauh lebih rendah dibandingkan dengan Generasi Boomers dan Generasi X. Generasi Z memiliki angka bunuh diri yang paling rendah yang mungkin disebabkan oleh tingginya kepedulian mental health pada gen Z. Beban pekerjaan yang berat dan tanggung jawab keluarga juga bisa menjadi faktor utama.

6. Bar plot banyaknya angka bunuh diri berdasarkan Country.



Dilakukan untuk melihat banyaknya angka bunuh diri berdasarkan perbedaan negara. Secara visual dapat dilihat dengan jelas bahwa United States memiliki tingkat bunuh diri yang paling tinggi disusul oleh Korea. Hal ini mungkin disebabkan oleh tekanan sosial yang tinggi (seperti kesenjangan sosial) di negara maju seperti Amerika dan Korea. Jumlah kasus untuk Korea = 96940 sementara Amerika = 292228.

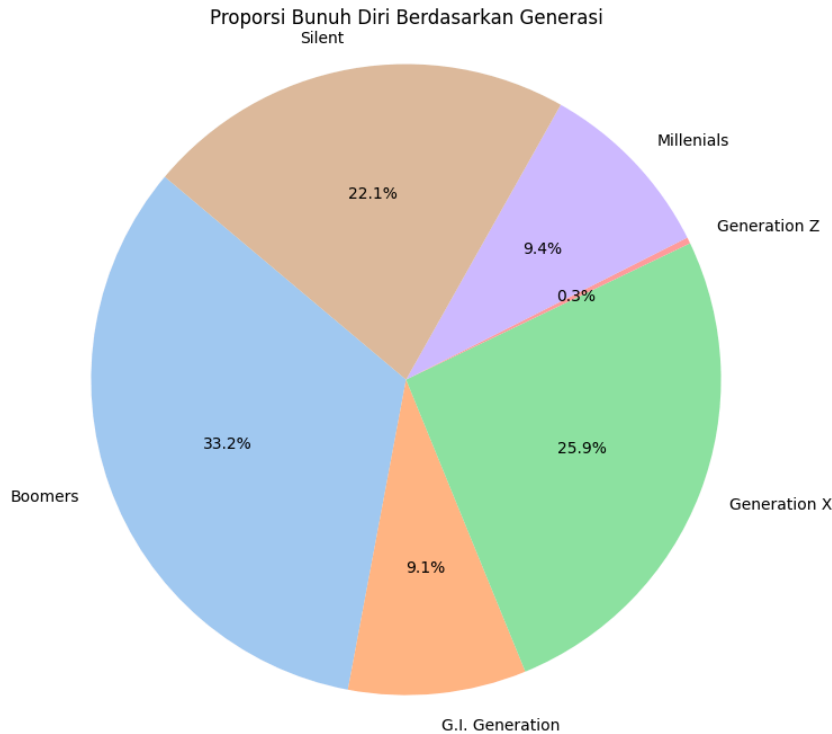
7. Heatmap untuk distribusi angka bunuh diri berdasarkan negara dan gender



Dari plot heatmap di atas kita dapat melihat distribusi pola bunuh diri di berbagai negara berdasarkan gender. Kasus tertinggi berada di negara United States dengan

gender pria diikuti kasus gender wanita di negara yang sama. Sedangkan kasus terendah berada di negara Kiribati dengan gender wanita. Visualisasi ini sangat berguna bagi pemerintah untuk melakukan tindak kebijakan sosial terhadap kasus bunuh diri di negaranya.

8. Pie chart untuk proporsi bunuh diri berdasarkan generasi



Hasil ini similar dengan visualisasi no.5 hanya saja disajikan dalam bentuk diagram lingkaran.

Bagian 3. Metode

Analisis Korespondensi Sederhana (*Simple Correspondence Analysis, CA*) adalah teknik statistik eksploratif yang digunakan untuk menganalisis hubungan antara dua variabel kategorikal. Teknik ini sangat berguna untuk mengidentifikasi pola dalam data yang disajikan dalam bentuk tabel kontingensi, yang menunjukkan frekuensi kejadian atau data jumlah dari kategori-kategori yang berbeda.

Prinsip Dasar

- Tabel Kontingensi: CA beroperasi pada tabel kontingensi yang menyajikan data dalam bentuk matriks, di mana baris dan kolom mewakili kategori dari variabel yang berbeda dan sel-sel matriks berisi frekuensi kejadian atau data numerik lainnya.
- Reduksi Dimensi: Tujuan utama CA adalah untuk mengurangi dimensi data, mengubahnya menjadi bentuk yang lebih sederhana yang mudah dipahami dan dianalisis. Ini dilakukan dengan menguraikan data asli ke dalam komponen-komponen yang menjelaskan sebagian besar variansi dalam data.

- Proyeksi Geometris: CA memproyeksikan data ke dalam ruang dua dimensi, di mana posisi relatif dari titik-titik (mewakili kategori) menggambarkan hubungan dan asosiasi antar kategori. Jarak antara titik-titik dalam ruang ini mengindikasikan kekuatan asosiasi antara kategori tersebut.

Langkah-Langkah Utama dalam CA

- Pembentukan Tabel Kontingensi: Data dikumpulkan dalam bentuk tabel kontingensi, di mana setiap sel berisi frekuensi atau jumlah kejadian untuk kombinasi kategori yang berbeda dari variabel yang dianalisis.

Contingency Table with a Rows and b Columns

		Columns				Row Total
		1	2	...	b	
Rows	1	n_{11}	n_{12}	...	n_{1b}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2b}	$n_{2.}$
	\vdots	\vdots	\vdots		\vdots	\vdots
	a	n_{a1}	n_{a2}	...	n_{ab}	$n_{a.}$
Column Total		$n_{.1}$	$n_{.2}$...	$n_{.b}$	n

$$\text{Row sum: } n_{i.} = \sum_{j=1}^b n_{ij}$$

$$\text{Total: } n = \sum_{ij} n_{ij}$$

$$\text{Column sum: } n_{.j} = \sum_{i=1}^a n_{ij}$$

$$i = 1, \dots, a$$

$$j = 1, \dots, b$$

Keterangan:

n_{ij} = jumlah pengamatan pada baris ke- i dan kolom ke- j

$n_{i.}$ = jumlah pengamatan pada baris ke- i

$n_{.j}$ = jumlah pengamatan pada kolom ke- j

n = total keseluruhan

- Normalisasi Data: Tabel kontingensi dinormalisasi untuk mengubah data mentah menjadi proporsi atau nilai yang dapat dibandingkan secara langsung yang disebut matriks korespondensi yang dinotasikan dengan **P**.

$$\mathbf{P} = (p_{ij}) = (n_{ij}/n)$$

Correspondence Matrix of Relative Frequencies

		Columns				Row Total
		1	2	...	b	
Rows	1	p_{11}	p_{12}	...	p_{1b}	$p_{1.}$
	2	p_{21}	p_{22}	...	p_{2b}	$p_{2.}$
	\vdots	\vdots	\vdots		\vdots	\vdots
	a	p_{a1}	p_{a2}	...	p_{ab}	$p_{a.}$
	Column Total	$p_{.1}$	$p_{.2}$...	$p_{.b}$	1

$$\text{Row sum: } p_{i.} = \sum_{j=1}^b p_{ij}$$

$$\text{Column sum: } p_{.j} = \sum_{i=1}^a p_{ij}$$

$$i = 1, \dots, a$$

$$j = 1, \dots, b$$

Keterangan:

p_{ij} = frekuensi relatif pada baris ke-i dan kolom ke-j

$p_{i.}$ = jumlah frekuensi relatif pada baris ke-i

$p_{.j}$ = jumlah frekuensi relatif pada kolom ke-j

Kita dapat menyusun vektor kolom yang dinotasikan dengan \mathbf{r} sebagai berikut:

$$\mathbf{r} = \mathbf{P}\mathbf{j} = (p_{1.}, p_{2.}, \dots, p_{a.})' = (n_{1.}/n, n_{2.}/n, \dots, n_{a.}/n)'$$

Dengan \mathbf{j} merupakan vektor $a \times 1$. Kita juga dapat menyusun vektor baris yang dinotasikan dengan \mathbf{c}' sebagai berikut:

$$\mathbf{c}' = \mathbf{j}'\mathbf{P} = (p_{.1}, p_{.2}, \dots, p_{.b}) = (n_{.1}/n, n_{.2}/n, \dots, n_{.b}/n)$$

Lalu kita dapat menyusun matriks \mathbf{R} dari row profiles yang diekspresikan dalam bentuk:

$$\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{P} = \begin{pmatrix} \mathbf{r}'_1 \\ \mathbf{r}'_2 \\ \vdots \\ \mathbf{r}'_a \end{pmatrix} = \begin{pmatrix} \frac{p_{11}}{p_{1.}} & \frac{p_{12}}{p_{1.}} & \dots & \frac{p_{1b}}{p_{1.}} \\ \frac{p_{21}}{p_{2.}} & \frac{p_{22}}{p_{2.}} & \dots & \frac{p_{2b}}{p_{2.}} \\ \vdots & \vdots & & \vdots \\ \frac{p_{a1}}{p_{a.}} & \frac{p_{a2}}{p_{a.}} & \dots & \frac{p_{ab}}{p_{a.}} \end{pmatrix}$$

$$\mathbf{D}_r = \text{diag}(\mathbf{r}) = \begin{pmatrix} p_{1.} & 0 & \dots & 0 \\ 0 & p_{2.} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & p_{a.} \end{pmatrix}$$

Dan matriks \mathbf{C} dari column profile yang diekspresikan dalam bentuk:

$$\mathbf{C} = \mathbf{P}\mathbf{D}_c^{-1} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_b) = \begin{pmatrix} \frac{p_{11}}{p_{.1}} & \frac{p_{12}}{p_{.2}} & \dots & \frac{p_{1a}}{p_{.a}} \\ \frac{p_{21}}{p_{.1}} & \frac{p_{22}}{p_{.2}} & \dots & \frac{p_{2a}}{p_{.a}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{a1}}{p_{.1}} & \frac{p_{a2}}{p_{.2}} & \dots & \frac{p_{ab}}{p_{.a}} \end{pmatrix}$$

$$\mathbf{D}_c = \text{diag}(\mathbf{c}) = \begin{pmatrix} p_{.1} & 0 & \dots & 0 \\ 0 & p_{.2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{.b} \end{pmatrix}$$

- Menguji independensi antar dua variabel kategorik menggunakan *chi-square test*

Hipotesis

H_0 : Tidak ada hubungan antar variabel A dan B

H_1 : Ada hubungan antar variabel A dan B

Statistik uji

$$\chi^2 = n \sum_{i=1}^a \sum_{j=1}^b \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}}$$

Derajat bebas: $df = (a - 1) \cdot (b - 1)$

Aturan penolakan

H_0 ditolak apabila nilai $\chi^2 > \chi_{\alpha, df}^2$ (nilai kritis) atau p-value $< \alpha$

- Perepresentasian Data dalam Ruang Geometris: Data dinormalisasi kemudian direpresentasikan dalam ruang geometris yang lebih rendah, biasanya dua dimensi, menggunakan dekomposisi nilai singular \mathbf{Z} (SVD) atau fungsi langsung di R.

$$\mathbf{Z} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_c^{-1/2}$$

$$\begin{aligned} \mathbf{Z}\mathbf{Z}' &= \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_c^{-1/2}\mathbf{D}_c^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')'\mathbf{D}_r^{-1/2} \\ &= \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_c^{-1}(\mathbf{P} - \mathbf{r}\mathbf{c}')'\mathbf{D}_r^{-1/2}. \end{aligned}$$

$$\mathbf{Z} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'$$

Keterangan:

\mathbf{U} = vektor eigen matriks $\mathbf{Z}\mathbf{Z}'$

$k = \min(a-1, b-1)$

$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$

\mathbf{V} = vektor eigen $\mathbf{Z}'\mathbf{Z}$

Elemen $\lambda_1, \lambda_2, \dots, \lambda_k$ dari matriks diagonal $\mathbf{\Lambda}$ disebut sebagai nilai singular dari \mathbf{Z} .

Untuk membuat plot koordinat deviasi row profile

$$\mathbf{X} = \mathbf{D}_r^{-1}\mathbf{A}\mathbf{\Lambda}.$$

$$\mathbf{X}_1 = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{a1} & x_{a2} \end{pmatrix}$$

$$\mathbf{A} = \mathbf{D}_r^{1/2} \mathbf{U}$$

Untuk membuat plot koordinat deviasi column profile

$$\mathbf{Y} = \mathbf{D}_c^{-1} \mathbf{B} \mathbf{A}.$$

$$\mathbf{Y}_1 = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{b1} & y_{b2} \end{pmatrix}$$

$$\mathbf{B} = \mathbf{D}_c^{1/2} \mathbf{V}$$

- Dekomposisi Inersia
Inersia total baris:

$$\text{Total inertia} = \frac{\chi^2}{n} = \sum_{i=1}^a p_{i.} (\mathbf{r}_i - \mathbf{c})' \mathbf{D}_c^{-1} (\mathbf{r}_i - \mathbf{c})$$

Inersia total kolom:

$$\text{Total inertia} = \frac{\chi^2}{n} = \sum_{j=1}^b p_{.j} (\mathbf{c}_j - \mathbf{r})' \mathbf{D}_r^{-1} (\mathbf{c}_j - \mathbf{r}) = \sum_{i=1}^k \lambda_i^2$$

- Jarak Profil
 - a. Jarak antara dua baris ke- i dan ke- i' adalah:

$$d^2(i, i') = \sum_{j=1}^b \frac{1}{p_{.j}} \left(\frac{p_{ij}}{p_{i.}} - \frac{p_{i'j}}{p_{i'.}} \right)^2$$

- b. Jarak antara dua kolom ke- j dan ke- j' adalah:

$$d^2(j, j') = \sum_{i=1}^a \frac{1}{p_{i.}} \left(\frac{p_{ij}}{p_{.j}} - \frac{p_{ij'}}{p_{.j'}} \right)^2$$

- Interpretasi Hasil: Plot dua dimensi dihasilkan untuk visualisasi hubungan antar kategori. Titik-titik yang berdekatan menunjukkan adanya asosiasi yang lebih kuat, sedangkan titik-titik yang berjauhan menunjukkan hubungan yang lebih lemah atau tidak ada hubungan.

Penelitian ini menggunakan Analisis Korespondensi Sederhana untuk mengeksplorasi dan memahami hubungan antara variabel-variabel kategorikal (negara, kelompok umur, dan generasi) terhadap tingkat bunuh diri (suicides_no). CA dipilih karena mampu memberikan visualisasi interaksi antar kategori dan mengungkap pola yang tersembunyi dalam data multidimensional.

Bagian 4. Pengolahan Data Dan Analisis Hasil

Pada penelitian ini akan dilakukan analisis korespondensi sederhana pada 3 variabel kategorik, yaitu variabel age, country dan generation terhadap angka bunuh diri. Proses analisis ini akan dibagi menjadi tiga bagian, yaitu analisis korespondensi variabel age dan country, variabel age dan generation, dan yang terakhir variabel generation dan country. Pada tahapan ini akan digunakan bantuan program R Studio.

Import Dataset Hasil Pre-Processing

```
#Import dataset bersih
df1 <- read.csv("C:\\Users\\Asus\\Downloads\\suicides1.csv")
View(df1)
```

Berikut ini cuplikan dari 10 observasi dataset bersih:

	country	sex	age	generation	suicides_no
1	United States	female	5-14 years	Generation Z	151
2	Guatemala	female	5-14 years	Generation X	2
3	Guatemala	male	5-14 years	Generation Z	6
4	Kiribati	female	5-14 years	Millenials	1
5	Guatemala	female	5-14 years	Millenials	4
6	United States	female	5-14 years	Generation X	52
7	United States	male	5-14 years	Generation X	195
8	Guatemala	male	5-14 years	Generation X	6
9	Kiribati	male	5-14 years	Millenials	1
10	Guatemala	female	5-14 years	Generation X	2

*Note: variabel sex hanya digunakan untuk kebutuhan visualisasi data agar menambah insight tetapi tidak akan digunakan dalam analisis korespondensi sederhana.

Analisis Korespondensi Sederhana untuk Variabel Age Dan Country

Berikut tahapan untuk analisis korespondensi sederhana dari variabel age dan country:

1. Membuat tabel kontingensi


```
> table_age_country
```

	5-14 years	15-24 years	25-34 years	35-54 years	55-74 years	75+ years
Albania	11	13	16	9	19	8
Guatemala	26	20	20	23	26	19
Kiribati	7	4	6	3	9	3
Republic of Korea	17	16	21	20	21	19
United Arab Emirates	2	4	6	3	5	4
United States	25	16	21	17	23	18

2. Menghitung matriks korespondensi P

```
> mat_korespondensi_age_country
```

	5-14 years	15-24 years	25-34 years	35-54 years	55-74 years	75+ years
Albania	0.14473684	0.17105263	0.21052632	0.11842105	0.25000000	0.10526316
Guatemala	0.19402985	0.14925373	0.14925373	0.17164179	0.19402985	0.14179104
Kiribati	0.21875000	0.12500000	0.18750000	0.09375000	0.28125000	0.09375000
Republic of Korea	0.14912281	0.14035088	0.18421053	0.17543860	0.18421053	0.16666667
United Arab Emirates	0.08333333	0.16666667	0.25000000	0.12500000	0.20833333	0.16666667
United States	0.20833333	0.13333333	0.17500000	0.14166667	0.19166667	0.15000000

3. Menghitung matriks jumlah baris dan kolom

```
> row_sum_age_country
```

	Albania	Guatemala	Kiribati	Republic of Korea	United Arab Emirates	United States
	76	134	32	114	24	120

```
> column_sum_age_country
```

	5-14 years	15-24 years	25-34 years	35-54 years	55-74 years	75+ years
	88	73	90	75	103	71

4. Menghitung matriks R dan C untuk row profiles dan column profiles

```
> (R <- solve(Dr) %*% mat_korespondensi_age_country)
```

	5-14 years	15-24 years	25-34 years	35-54 years	55-74 years	75+ years
[1,]	0.001904432	0.002250693	0.002770083	0.001558172	0.003289474	0.001385042
[2,]	0.001447984	0.001113834	0.001113834	0.001280909	0.001447984	0.001058142
[3,]	0.006835938	0.003906250	0.005859375	0.002929688	0.008789062	0.002929688
[4,]	0.001308095	0.001231148	0.001615882	0.001538935	0.001615882	0.001461988
[5,]	0.003472222	0.006944444	0.010416667	0.005208333	0.008680556	0.006944444
[6,]	0.001736111	0.001111111	0.001458333	0.001180556	0.001597222	0.001250000

```
> (C <- mat_korespondensi_age_country %*% solve(Dc))
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
Albania	0.0016447368	0.002343187	0.002339181	0.001578947	0.002427184	0.001482580
Guatemala	0.0022048847	0.002044572	0.001658375	0.002288557	0.001883785	0.001997057
Kiribati	0.0024857955	0.001712329	0.002083333	0.001250000	0.002730583	0.001320423
Republic of Korea	0.0016945774	0.001922615	0.002046784	0.002339181	0.001788452	0.002347418
United Arab Emirates	0.0009469697	0.002283105	0.002777778	0.001666667	0.002022654	0.002347418
United States	0.0023674242	0.001826484	0.001944444	0.001888889	0.001860841	0.002112676

5. Menguji independensi antar variabel

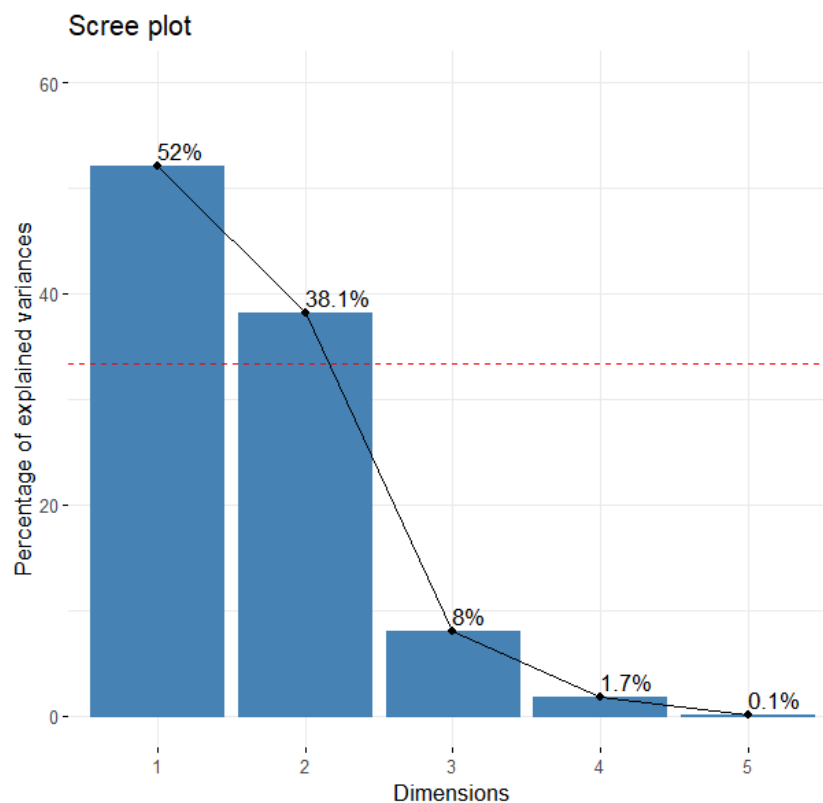
Pearson's Chi-squared test

```
data: table_age_country
X-squared = 12.011, df = 25, p-value = 0.9865
```

Karena $p\text{-value} > \alpha = 0.05$ maka H_0 tidak ditolak. Disimpulkan bahwa tidak ada hubungan yang signifikan antara variabel age dan country.

Meskipun tidak ada hubungan signifikan yang ditemukan melalui uji Chi-Square, analisis korespondensi dapat menunjukkan wawasan yang lebih mendalam tentang data dan membantu mengidentifikasi pola yang mungkin terlewatkan oleh analisis statistik sederhana.

6. Menentukan dimensi dari Scree plot



Disini kita akan mengambil persentase dengan minimal 80% menjelaskan data. Berdasarkan plot di atas hanya dimensi 1 (menjelaskan 52% dari total inersia) dan 2 (menjelaskan 38,1% dari total inersia) yang boleh digunakan dalam penyelesaiannya. Dimensi 3, 4 dan 5 menjelaskan total inersia di bawah rata-rata eigenvalue (garis putus-putus merah).

7. Matriks X1 (row profiles) dan Y1 (column profiles) berturut-turut

```
> ca$row$coord[, 1:2]
```

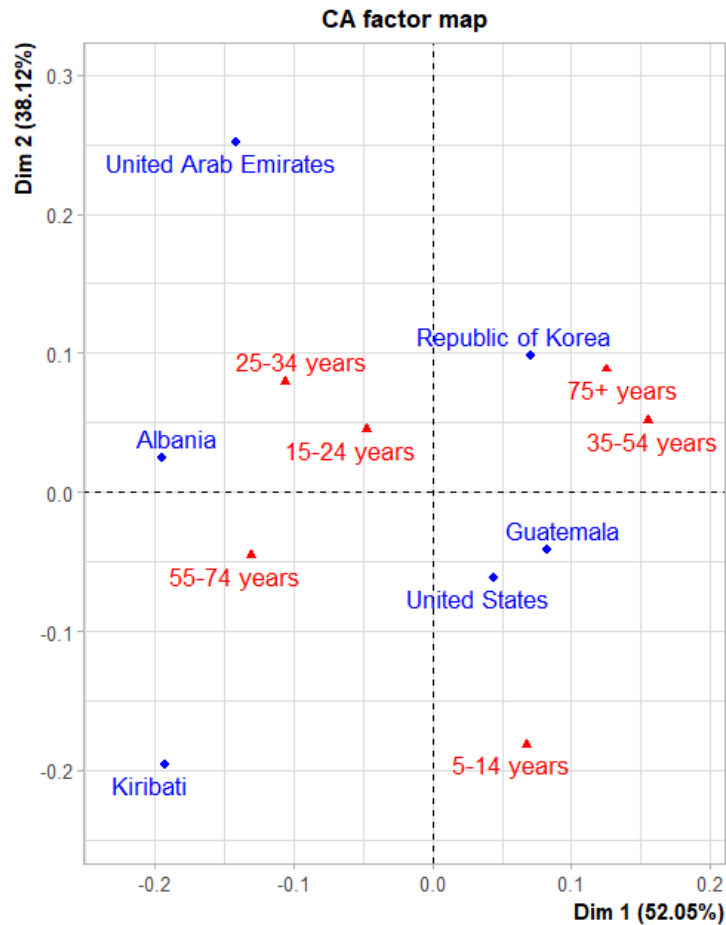
	Dim 1	Dim 2
Albania	-0.19487382	0.02441646
Guatemala	0.08243001	-0.04100665
Kiribati	-0.19202153	-0.19540874
Republic of Korea	0.07037374	0.09866519
United Arab Emirates	-0.14116460	0.25166632
United States	0.04395686	-0.06162921

```
> ca$col$coord[, 1:2]
```

	Dim 1	Dim 2
5-14 years	0.06792518	-0.18116795
15-24 years	-0.04756088	0.04507877
25-34 years	-0.10608373	0.07963700
35-54 years	0.15468547	0.05169204
55-74 years	-0.13064111	-0.04546602
75+ years	0.12530548	0.08860277

8. Plot koordinat row dan column profile dengan fungsi langsung di R yaitu 'CA'

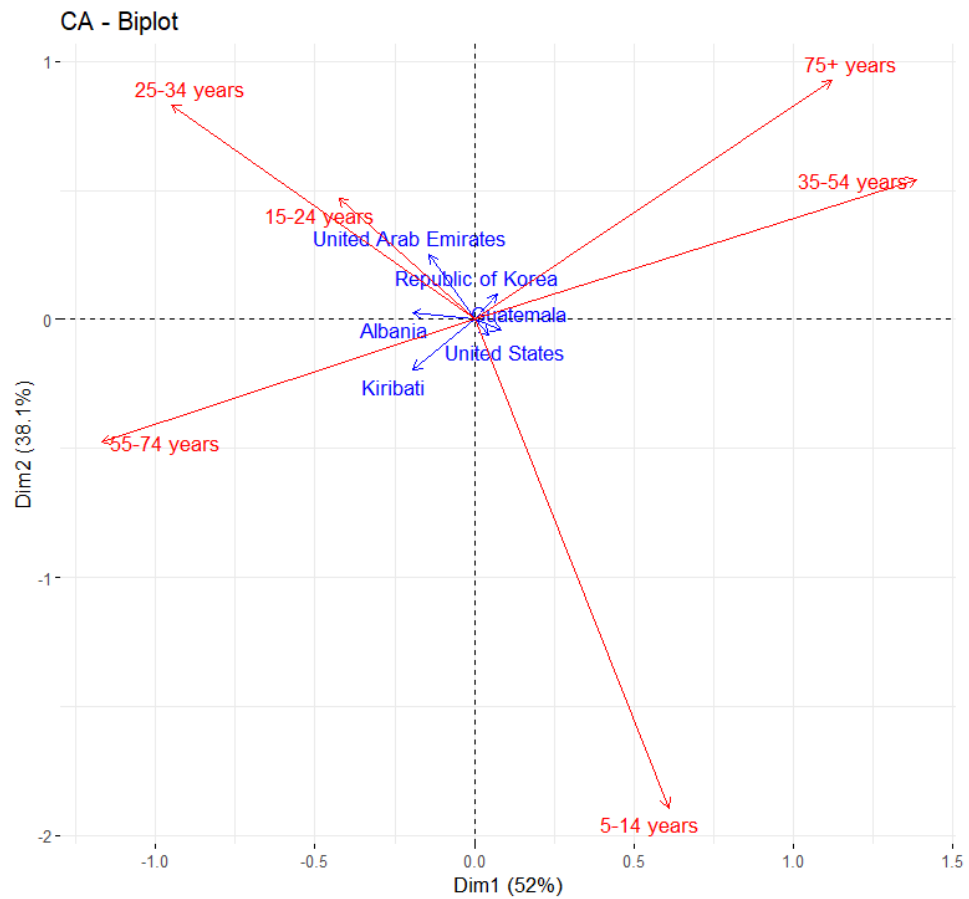
```
> ca <- CA(table_age_country, graph = TRUE)
```



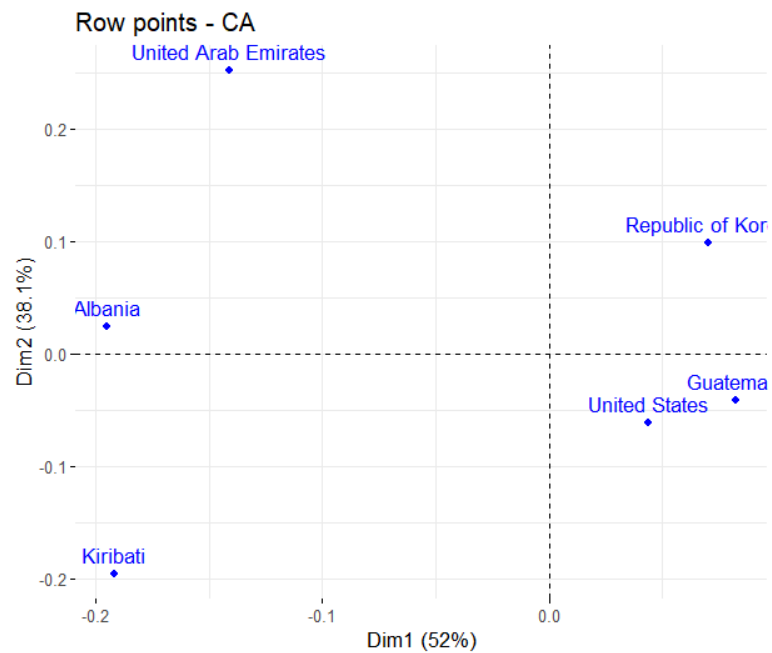
Interpretasi biplot Age dan Country:

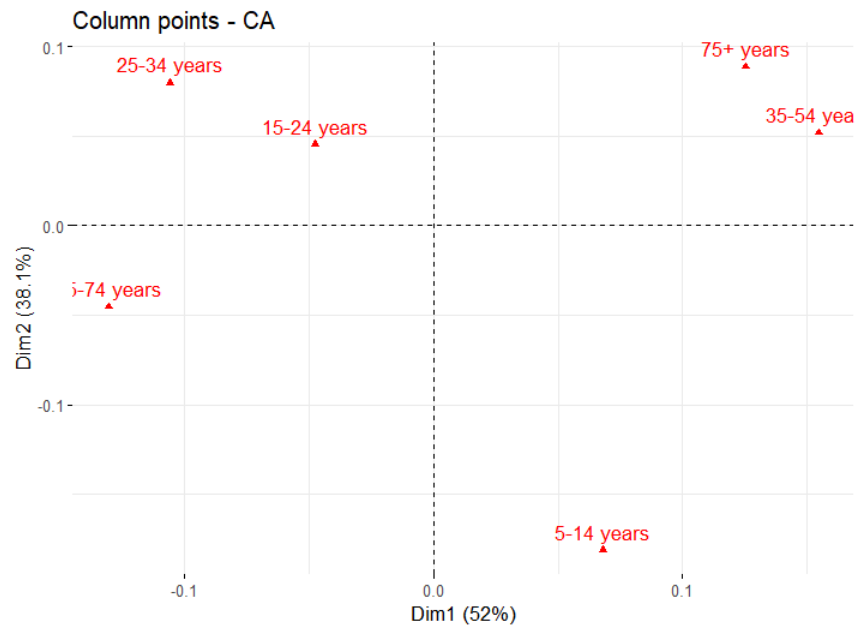
- Dimensi 1 (52.05%): Dimensi pertama menjelaskan 55.74% dari total variabilitas dalam data.
- Dimensi 2 (38.12%): Dimensi kedua menjelaskan 29.68% dari total variabilitas dalam data.
- Guatemala dan United States: Memiliki hubungan kuat dengan kasus bunuh diri pada anak-anak (5-14 tahun).
- Albania dan Arab: Menunjukkan angka kasus bunuh diri yang lebih tinggi pada remaja (15-24 tahun) dan orang dewasa (55-74 tahun).
- Republic of Korea: Menunjukkan angka kasus bunuh diri yang tinggi pada dewasa paruh baya (35-54 tahun) dan lansia (75+ tahun).
- Kiribati: Memiliki pola kasus bunuh diri yang berbeda atau lebih rendah secara keseluruhan.

Biplot asimetris



Visualisasi row profile dan column profiles secara terpisah





9. Plot kontribusi

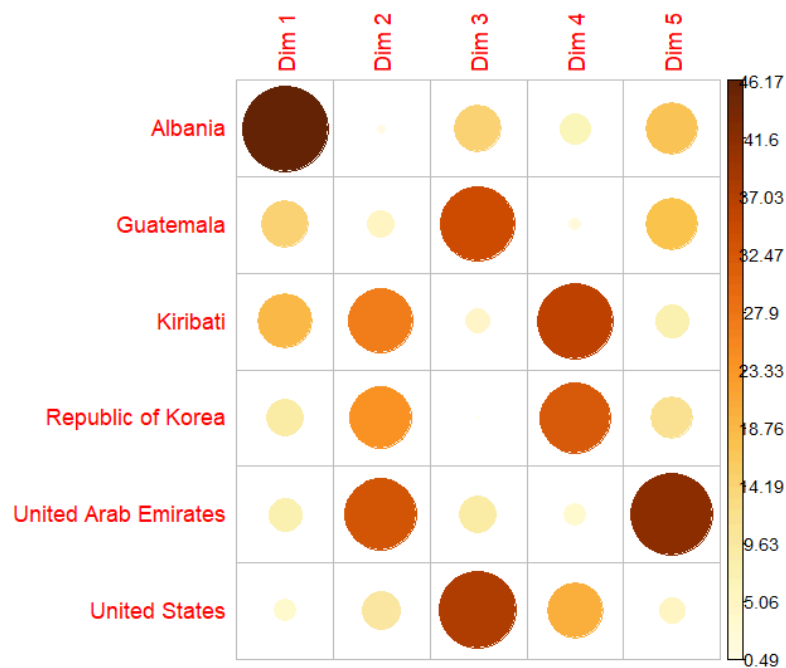
Baris

Kontribusi baris dalam persen

```
> head(row$contrib)
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Albania	46.169210	0.9896665	14.2613280	6.370882	17.008913
Guatemala	14.564894	4.9217893	34.4842789	1.535014	17.694023
Kiribati	18.874770	26.6899420	4.1944698	36.342741	7.498077
Republic of Korea	9.031459	24.2405513	0.4890156	32.148195	11.290779
United Arab Emirates	7.650585	33.2025274	8.9346249	3.700829	41.711434
United States	3.709081	9.9555234	37.6362828	19.902338	4.796774

Corrplot

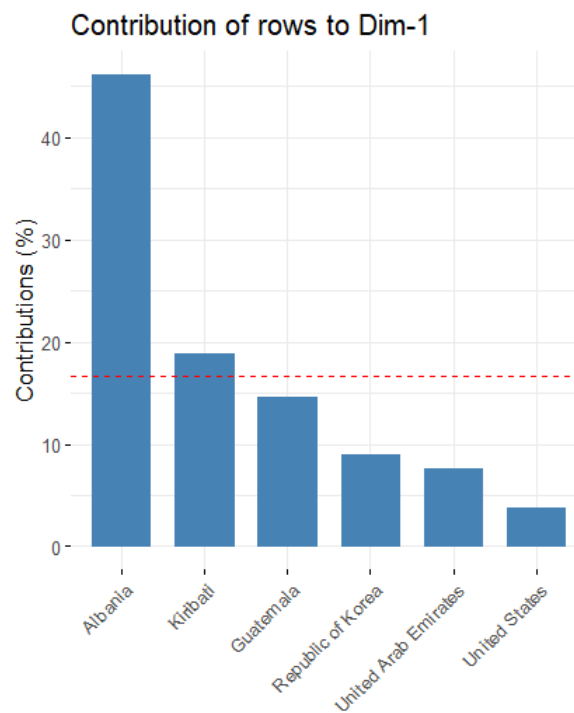


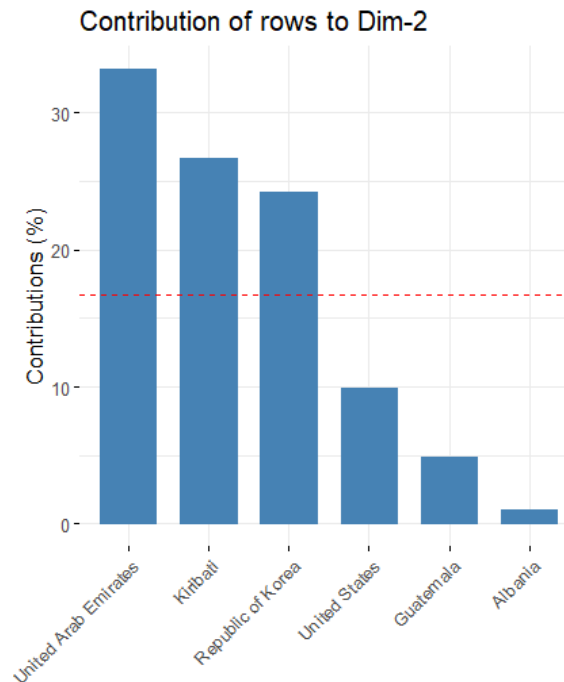
- **Albania:** Berkontribusi signifikan di Dimensi 1.
- **Guatemala:** Kontribusi besar di Dimensi 3.

- **Kiribati**: Kontribusi signifikan di Dimensi 4.
- **Republic of Korea**: Kontribusi besar di Dimensi 4.
- **United Arab Emirates**: Kontribusi terbesar di Dimensi 5.
- **United States**: Kontribusi signifikan di Dimensi 3.

Kontribusi di sini mengacu pada seberapa besar setiap dimensi dalam PCA menjelaskan variabilitas data kasus bunuh diri untuk setiap negara. Dimensi dengan kontribusi lebih besar menunjukkan bahwa dimensi tersebut penting dalam menggambarkan pola kasus bunuh diri di negara tersebut.

Berikut ini adalah bar plot untuk melihat kontribusi dimensi 1 dan 2 dari baris





Dim-1:

- Albania memiliki kontribusi terbesar, diikuti oleh Kiribati dan Guatemala.
- Negara-negara ini mungkin memiliki karakteristik atau pola unik terkait kasus bunuh diri yang sangat mempengaruhi variasi pada dimensi pertama.

Dim-2:

- United Arab Emirates memiliki kontribusi terbesar, diikuti oleh Kiribati dan Republic of Korea.
- Negara-negara ini mungkin memiliki faktor-faktor unik atau pola yang berbeda dari yang terlihat di Dim-1 yang mempengaruhi kasus bunuh diri.

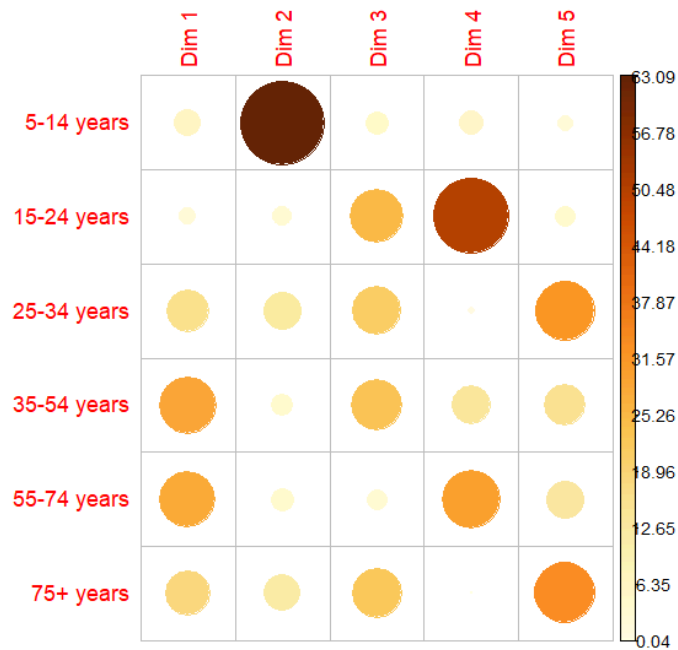
Kolom

Kontribusi kolom dalam persen

```
> ca$col$contrib
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
5-14 years	6.494954	63.089216	4.947576	5.49411560	2.374139
15-24 years	2.641523	3.240236	25.224147	50.36619695	3.927896
25-34 years	16.202124	12.467576	20.729338	0.49388638	32.107075
35-54 years	28.707285	4.377422	22.992125	13.61239129	15.310776
55-74 years	28.120871	4.650727	3.696024	29.98972579	12.942651
75+ years	17.833242	12.174823	22.410790	0.04368399	33.337461

Corrplot



Dim-1:

- Usia 5-14 tahun memiliki kontribusi terbesar terhadap Dim-1.
- Kelompok usia 35-54 years juga memiliki kontribusi besar.

Dim-2:

- Kelompok usia 5-14 years memiliki kontribusi besar.
- Ini mungkin mengindikasikan bahwa di United Arab Emirates, kasus bunuh diri pada kelompok usia 5-14 tahun sangat signifikan dalam menjelaskan variasi dalam Dim-2.

Dim-3:

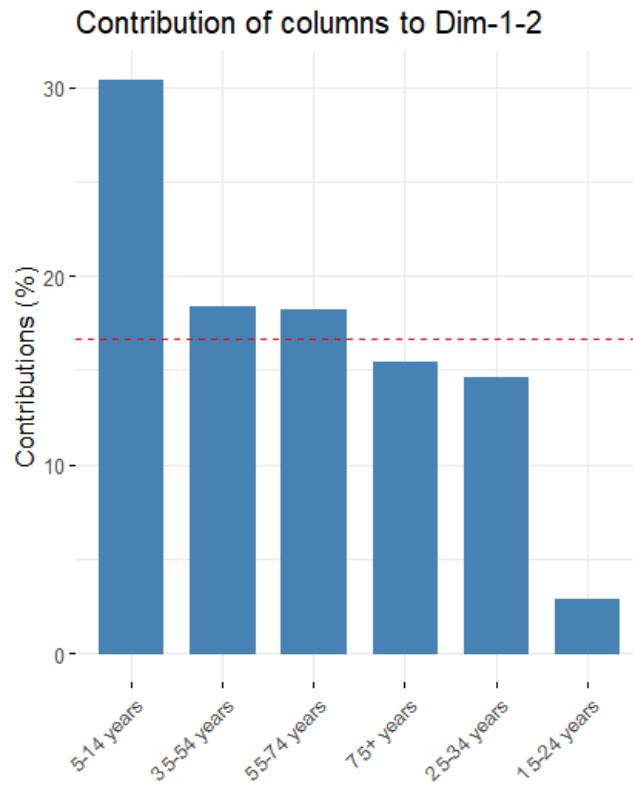
- 15-24 years memiliki kontribusi terbesar terhadap Dim-3.

Dim-4:

- 15-24 years juga menunjukkan kontribusi tinggi.

Dim-5:

- 25-34 years dan 75+ years menunjukkan kontribusi tinggi.



Terlihat bahwa kelompok umur 5-14 tahun memberikan kontribusi paling besar untuk kedua dimensi.

Analisis Korespondensi Sederhana untuk Variabel Age Dan Generation

Berikut tahapan untuk analisis korespondensi sederhana dari variabel age dan generation:

1. Tabel kontingensi

```
> print(table_age_generation)
```

	G.I. Generation	Silent	Boomers	Generation X	Millenials	Generation Z
5-14 years	0	0	0	17	51	20
15-24 years	0	0	0	36	37	0
25-34 years	0	0	29	49	12	0
35-54 years	0	12	45	18	0	0
55-74 years	15	72	16	0	0	0
75+ years	32	39	0	0	0	0

2. Matriks korespondensi

```
> mat_korespondensi_age_generation
```

	G.I. Generation	Silent	Boomers	Generation X	Millenials	Generation Z
5-14 years	0.0000000	0.0000000	0.0000000	0.1931818	0.5795455	0.2272727
15-24 years	0.0000000	0.0000000	0.0000000	0.4931507	0.5068493	0.0000000
25-34 years	0.0000000	0.0000000	0.3222222	0.5444444	0.1333333	0.0000000
35-54 years	0.0000000	0.1600000	0.6000000	0.2400000	0.0000000	0.0000000
55-74 years	0.1456311	0.6990291	0.1553398	0.0000000	0.0000000	0.0000000
75+ years	0.4507042	0.5492958	0.0000000	0.0000000	0.0000000	0.0000000

3. Row sum dan column sum

```
> row_sum_age_generation
```

5-14 years	15-24 years	25-34 years	35-54 years	55-74 years	75+ years
88	73	90	75	103	71

```
> column_sum_age_generation
G.I. Generation      Silent      Boomers      Generation X      Millenials      Generation Z
      47             123             90             120             100             20
```

4. Menghitung matriks **R** dan **C** untuk row profiles dan column profiles

R =

```
      G.I. Generation      Silent      Boomers      Generation X      Millenials      Generation Z
[1,] 0.000000000 0.000000000 0.000000000 0.002195248 0.006585744 0.002582645
[2,] 0.000000000 0.000000000 0.000000000 0.006755489 0.006943141 0.000000000
[3,] 0.000000000 0.000000000 0.003580247 0.006049383 0.001481481 0.000000000
[4,] 0.000000000 0.002133333 0.008000000 0.003200000 0.000000000 0.000000000
[5,] 0.001413894 0.006786691 0.001508153 0.000000000 0.000000000 0.000000000
[6,] 0.006347947 0.007736560 0.000000000 0.000000000 0.000000000 0.000000000
```

C =

```
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
5-14 years 0.000000000 0.000000000 0.000000000 0.001609848 0.005795455 0.01136364
15-24 years 0.000000000 0.000000000 0.000000000 0.004109589 0.005068493 0.000000000
25-34 years 0.000000000 0.000000000 0.003580247 0.004537037 0.001333333 0.000000000
35-54 years 0.000000000 0.001300813 0.006666667 0.002000000 0.000000000 0.000000000
55-74 years 0.003098533 0.005683164 0.001725998 0.000000000 0.000000000 0.000000000
75+ years  0.009589452 0.004465819 0.000000000 0.000000000 0.000000000 0.000000000
```

5. Menguji independensi antar variabel

```
> chisq.test(table_age_generation)
```

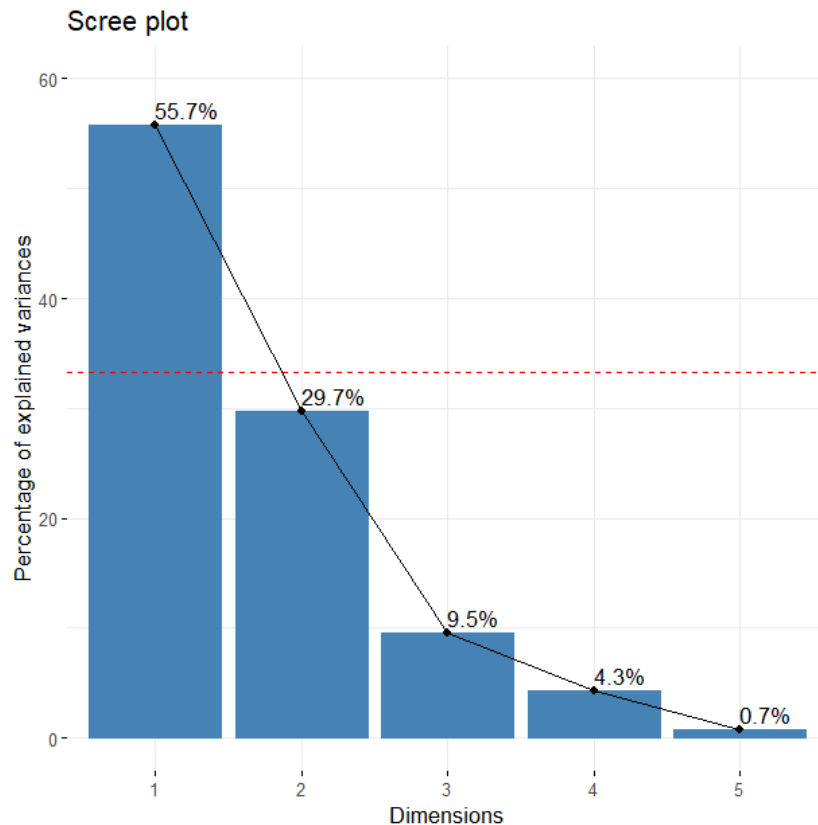
Pearson's Chi-squared test

data: table_age_generation

X-squared = 771.88, df = 25, p-value < 2.2e-16

Karena $p\text{-value} < \alpha = 0.05$ maka H_0 ditolak. Disimpulkan bahwa ada hubungan yang signifikan antara variabel age dan generation.

6. Menentukan dimensi dari Scree plot



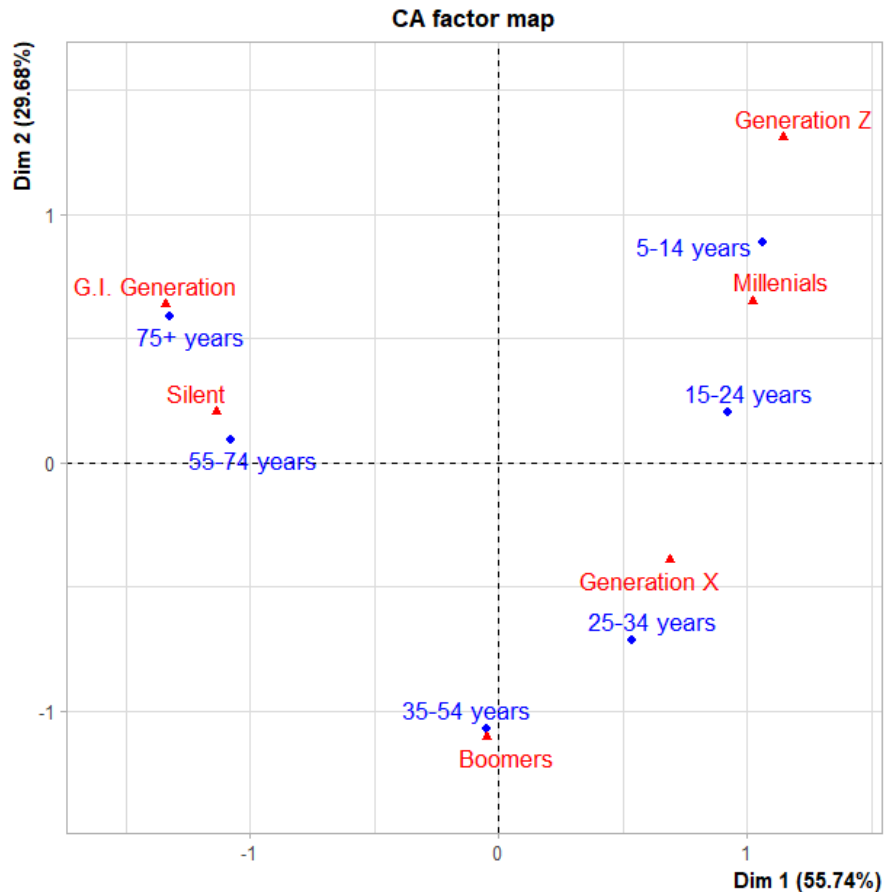
Disini kita akan mengambil persentase dengan minimal 80% menjelaskan data. Berdasarkan plot di atas hanya dimensi 1 (menjelaskan 55.7% dari total inersia) yang boleh digunakan dalam penyelesaiannya. Dimensi 2, 3, 4 dan 5 menjelaskan total inersia di bawah rata-rata eigenvalue (garis putus-putus merah).

- Matriks X1 (row profiles) dan Y1 (column profiles) berturut-turut

```
> ca$row$coord[, 1:2]
      Dim 1      Dim 2
5-14 years  1.05942136  0.88545558
15-24 years  0.92178621  0.20275556
25-34 years  0.53301699 -0.71224829
35-54 years -0.04895547 -1.06958683
55-74 years -1.07505381  0.09419745
75+ years  -1.32519435  0.59010951
> ca$col$coord[, 1:2]
      Dim 1      Dim 2
G.I. Generation -1.34248279  0.6379735
Silent          -1.13647674  0.2037219
Boomers         -0.04726805 -1.1043836
Generation X     0.68659699 -0.3915044
Millenials       1.01904994  0.6517015
Generation Z     1.14204115  1.3081177
```

- Plot koordinat row dan column profile dengan fungsi langsung di R yaitu 'CA'

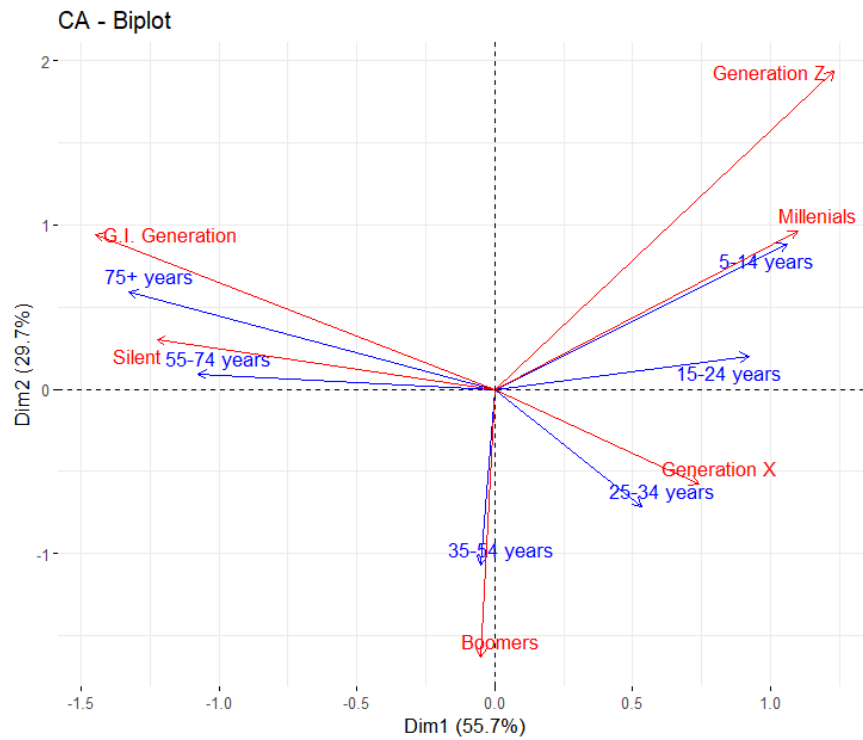
```
> ca <- CA(table_age_generation, graph = TRUE)
```



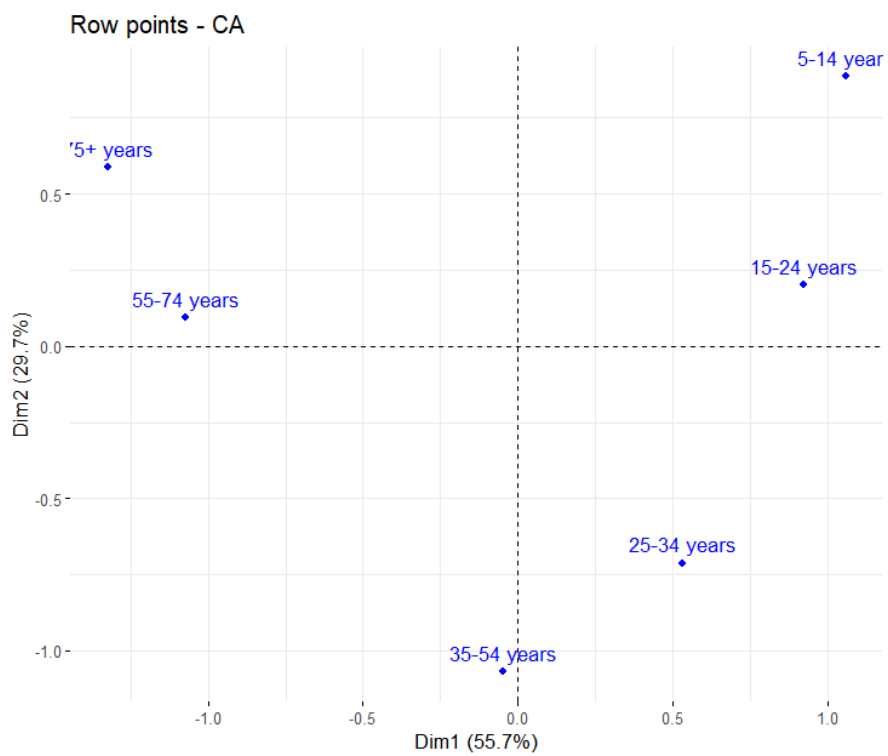
Interpretasi biplot Age dan Generation

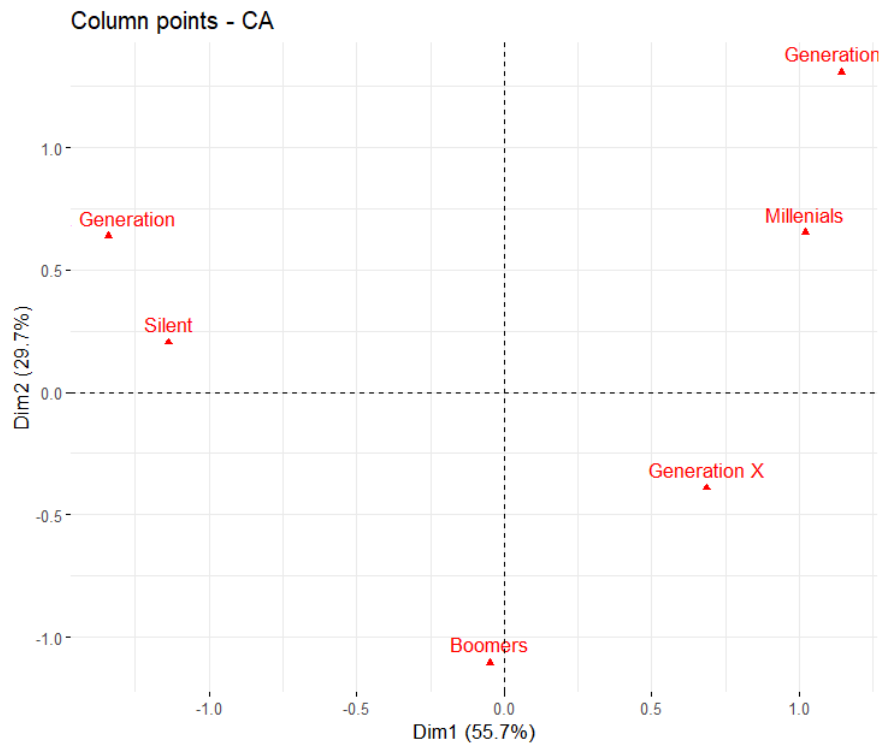
- Dimensi 1 (55.74%): Dimensi pertama menjelaskan 55.74% dari total variabilitas dalam data.
- Dimensi 2 (29.68%): Dimensi kedua menjelaskan 29.68% dari total variabilitas dalam data.
- Generasi Z (5-14 tahun): Ada perhatian khusus pada bunuh diri di usia muda, menunjukkan perlunya intervensi dini untuk masalah kesehatan mental dan tekanan sosial.
- Millenials (15-24 tahun): Kasus bunuh diri banyak terjadi di usia remaja akhir hingga dewasa muda, yang mungkin disebabkan oleh tekanan akademik dan masalah kesehatan mental.
- Generation X (25-34 tahun): Kasus bunuh diri pada usia dewasa muda ini terkait dengan tekanan karier dan masalah keluarga.
- Boomers (35-54 tahun): Risiko bunuh diri meningkat di usia pertengahan karena tekanan pekerjaan dan masalah kesehatan.
- Silent Generation dan G.I. Generation (55+ tahun): Kasus bunuh diri di usia tua terkait dengan isolasi sosial dan masalah kesehatan.

Biplot asimetris



Visualisasi row profile dan column profiles secara terpisah





9. Plot kontribusi

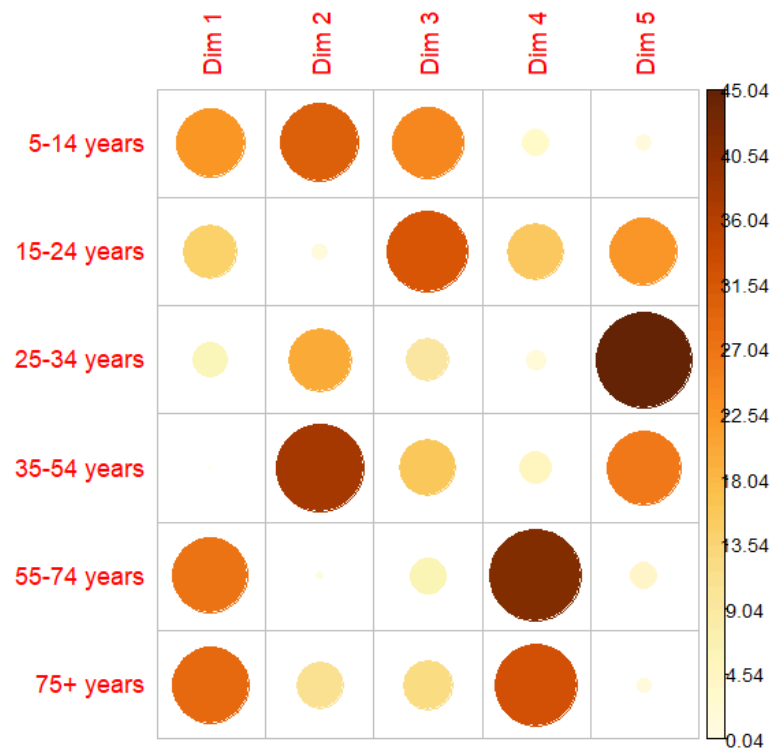
Baris

Kontribusi baris dalam persen

```
> head(row$contrib)
```

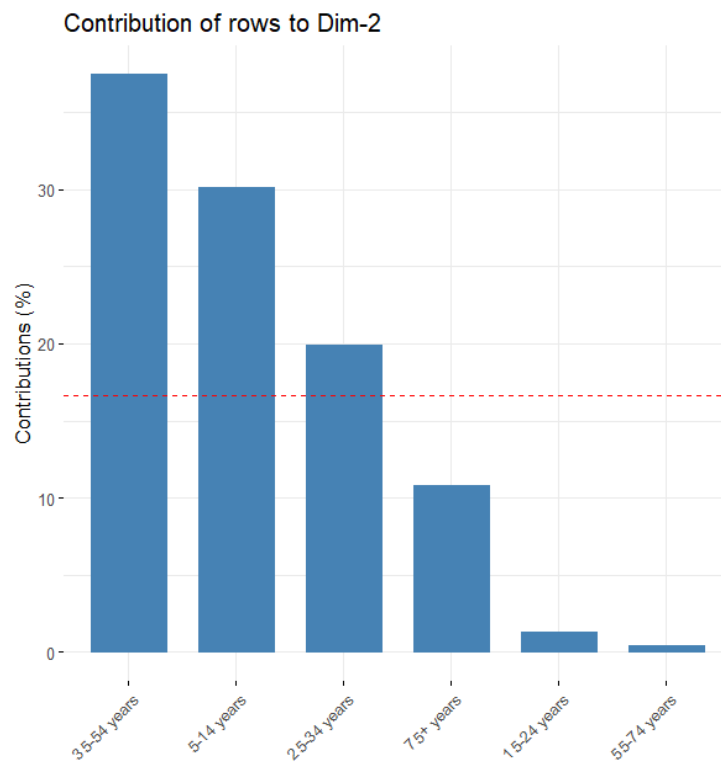
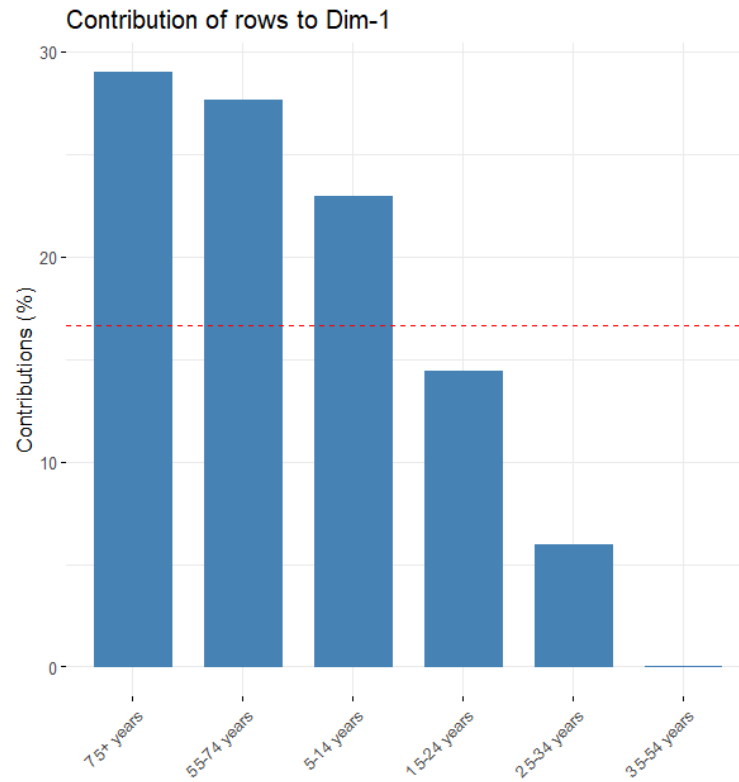
	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
5-14 years	22.95494051	30.1166260	24.664033	3.441770	1.222631
15-24 years	14.41581890	1.3099614	31.712794	15.420397	22.541029
25-34 years	5.94265684	19.9294555	9.122035	1.962925	45.042928
35-54 years	0.04177531	37.4527274	15.764848	5.323944	26.416706
55-74 years	27.66646543	0.3989382	6.311061	41.285642	3.737894
75+ years	28.97834301	10.7922915	12.425230	32.565323	1.038813

Corrplot



- **5-14:** Berkontribusi signifikan di Dimensi 2.
- **15-24:** Kontribusi besar di Dimensi 3.
- **25-34:** Kontribusi signifikan di Dimensi 5.
- **35-54:** Kontribusi besar di Dimensi 2.
- **55-74:** Kontribusi terbesar di Dimensi 4.
- **75+ :** Kontribusi signifikan di Dimensi 4.

Berikut bar plot untuk melihat kontribusi baris pada dimensi 1 dan 2



Dim-1:

Kelompok umur 75+ memiliki kontribusi terbesar, diikuti oleh 55-74 dan 5-14 tahun.

Dim-2:

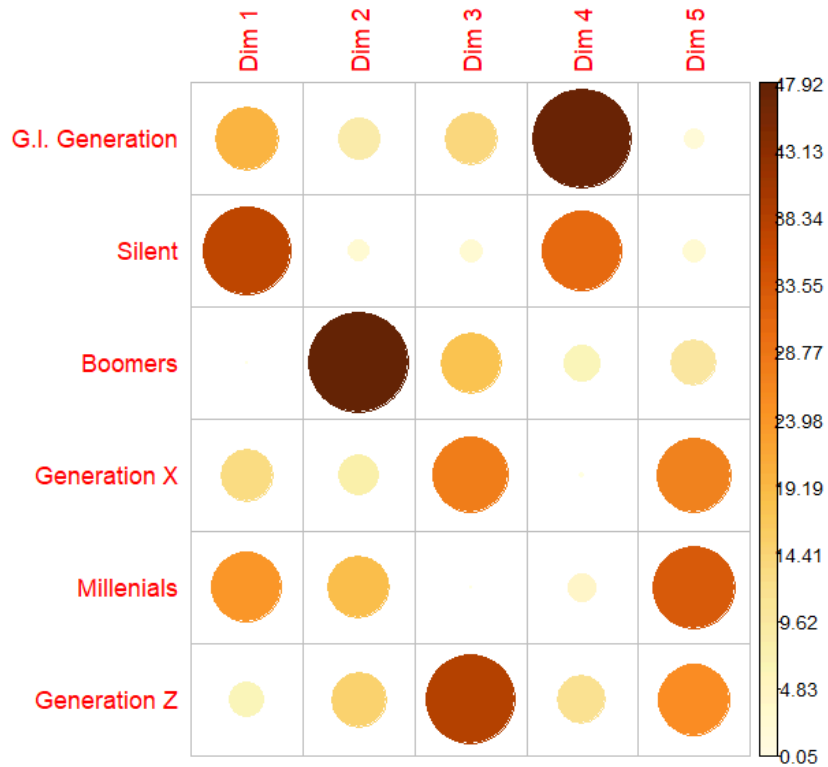
Kelompok umur 35-54 tahun memiliki kontribusi terbesar, diikuti oleh 5-14 dan 25-34 tahun.

Kolom

Kontribusi kolom dalam persen

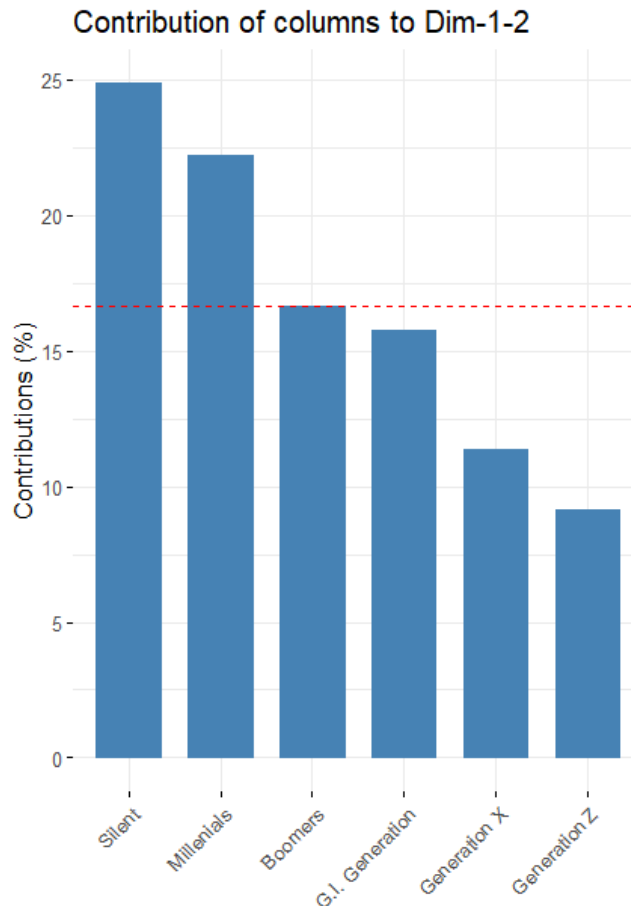
```
> head(col$contrib)
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
G.I. Generation	19.68662915	8.350129	13.55145836	47.0322814	1.979502
Silent	36.92174889	2.228285	2.71836675	30.9363164	2.595283
Boomers	0.04673411	47.915114	17.84010403	6.2513624	9.946686
Generation X	13.14743635	8.028688	27.63941937	0.1647984	27.019657
Millenials	24.13498229	18.539052	0.09923791	3.9507830	33.275945
Generation Z	6.06246921	14.938732	38.15141359	11.6644584	25.182927



- **GI Generation:** Berkontribusi signifikan di Dimensi 4.
- **Silent:** Kontribusi besar di Dimensi 3.
- **Boomers:** Kontribusi signifikan di Dimensi 2.
- **Gen X:** Kontribusi besar di Dimensi 3.
- **Millenials:** Kontribusi terbesar di Dimensi 5.
- **Gen Z :** Kontribusi signifikan di Dimensi 3.

Berikut ini adalah barplot gabungan untuk melihat kontribusi kolom di dimensi 1 dan 2.



Terlihat bahwa generasi Silent memberikan kontribusi paling besar di kedua dimensi diikuti oleh generasi milenial.

Analisis Korespondensi Sederhana untuk Variabel Generation Dan Country

Berikut tahapan untuk analisis korespondensi sederhana dari variabel generation dan country:

1. Tabel kontingensi

```
> print(table_generation_country)
```

	G.I. Generation	Silent	Boomers	Generation X	Millennials	Generation Z
Albania	4	25	11	18	16	2
Guatemala	14	28	27	32	28	5
Kiribati	3	9	6	7	7	0
Republic of Korea	12	27	23	23	23	6
United Arab Emirates	0	9	3	6	5	1
United States	14	25	20	34	21	6

2. Matriks korespondensi

```
> mat_korespondensi_generation_country
```

	G.I. Generation	Silent	Boomers	Generation X	Millennials	Generation Z
Albania	0.05263158	0.32894737	0.14473684	0.23684211	0.21052632	0.02631579
Guatemala	0.10447761	0.20895522	0.20149254	0.23880597	0.20895522	0.03731343
Kiribati	0.09375000	0.28125000	0.18750000	0.21875000	0.21875000	0.00000000
Republic of Korea	0.10526316	0.23684211	0.20175439	0.20175439	0.20175439	0.05263158
United Arab Emirates	0.00000000	0.37500000	0.12500000	0.25000000	0.20833333	0.04166667
United States	0.11666667	0.20833333	0.16666667	0.28333333	0.17500000	0.05000000

3. Row sum dan column sum

```
> row_sum_generation_country
```

	Albania	Guatemala	Kiribati	Republic of Korea	United Arab Emirates	United States
	76	134	32	114	24	120

```
> column_sum_generation_country
```

	G.I. Generation	Silent	Boomers	Generation X	Millennials	Generation Z
	47	123	90	120	100	20

4. Menghitung matriks **R** dan **C** untuk row profiles dan column profiles

R =

```
> (R <- solve(Dr) %*% mat_korespondensi_generation_country)
```

	G.I. Generation	Silent	Boomers	Generation X	Millennials	Generation Z
[1,]	0.0006925208	0.004328255	0.001904432	0.003116343	0.002770083	0.0003462604
[2,]	0.0007796837	0.001559367	0.001503676	0.001782134	0.001559367	0.0002784585
[3,]	0.0029296875	0.008789062	0.005859375	0.006835938	0.006835938	0.0000000000
[4,]	0.0009233610	0.002077562	0.001769775	0.001769775	0.001769775	0.0004616805
[5,]	0.0000000000	0.015625000	0.005208333	0.010416667	0.008680556	0.0017361111
[6,]	0.0009722222	0.001736111	0.001388889	0.002361111	0.001458333	0.0004166667

C =

```
> (C <- mat_korespondensi_generation_country %*% solve(Dc))
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
Albania	0.001119821	0.002674369	0.001608187	0.001973684	0.002105263	0.001315789
Guatemala	0.002222928	0.001698823	0.002238806	0.001990050	0.002089552	0.001865672
Kiribati	0.001994681	0.002286585	0.002083333	0.001822917	0.002187500	0.000000000
Republic of Korea	0.002239642	0.001925546	0.002241715	0.001681287	0.002017544	0.002631579
United Arab Emirates	0.000000000	0.003048780	0.001388889	0.002083333	0.002083333	0.002083333
United States	0.002482270	0.001693767	0.001851852	0.002361111	0.001750000	0.002500000

5. Menguji independensi antar variabel

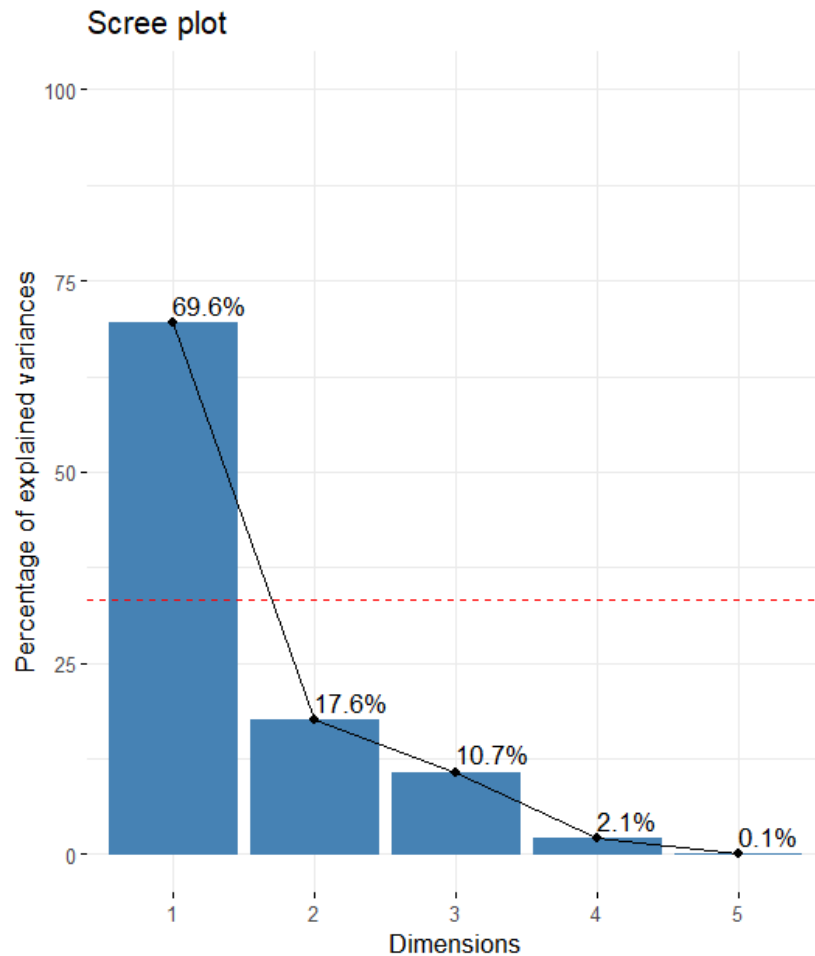
Pearson's Chi-squared test

```
data: table_generation_country
X-squared = 16.357, df = 25, p-value = 0.9038
```

Karena $p\text{-value} > \alpha = 0.05$ maka H_0 tidak ditolak. Disimpulkan bahwa tidak ada hubungan yang signifikan antara variabel generation dan country.

Meskipun tidak ada hubungan signifikan yang ditemukan melalui uji Chi-Square, analisis korespondensi dapat menunjukkan wawasan yang lebih mendalam tentang data dan membantu mengidentifikasi pola yang mungkin terlewatkan oleh analisis statistik sederhana.

6. Menentukan dimensi dari Scree plot



Disini kita akan mengambil persentase dengan minimal 80% menjelaskan data. Berdasarkan plot di atas hanya dimensi 1 (menjelaskan 69.6% dari total inersia) yang boleh digunakan dalam penyelesaiannya. Dimensi 2, 3, 4 dan 5 menjelaskan total inersia di bawah rata-rata eigenvalue (garis putus-putus merah).

7. Matriks X1 (row profiles) dan Y1 (column profiles) berturut-turut

```
> ca$row$coord[, 1:2]
```

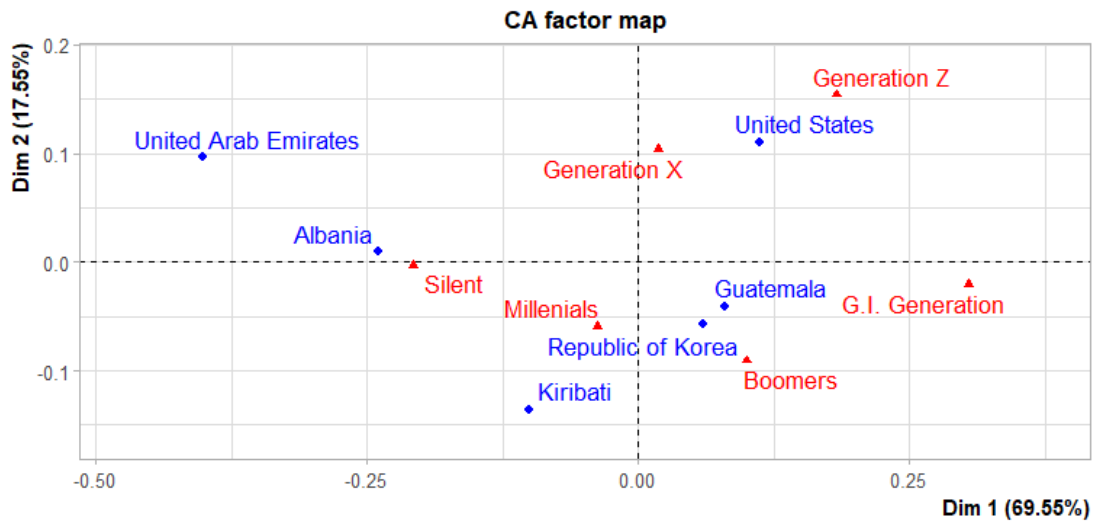
	Dim 1	Dim 2
Albania	-0.24027148	0.009283975
Guatemala	0.08055702	-0.040886358
Kiribati	-0.09988383	-0.135543798
Republic of Korea	0.05983531	-0.056643205
United Arab Emirates	-0.40189037	0.096057472
United States	0.11238682	0.110521145

```
> #Matriks Y (column profiles)
> ca$col$coord[, 1:2]
```

	Dim 1	Dim 2
G.I. Generation	0.30446221	-0.02086802
Silent	-0.20707327	-0.00369410
Boomers	0.09951783	-0.09079666
Generation X	0.01877096	0.10349967
Millenials	-0.03719471	-0.05893778
Generation Z	0.18353191	0.15403443

8. Plot koordinat row dan column profile dengan fungsi langsung di R yaitu 'CA'

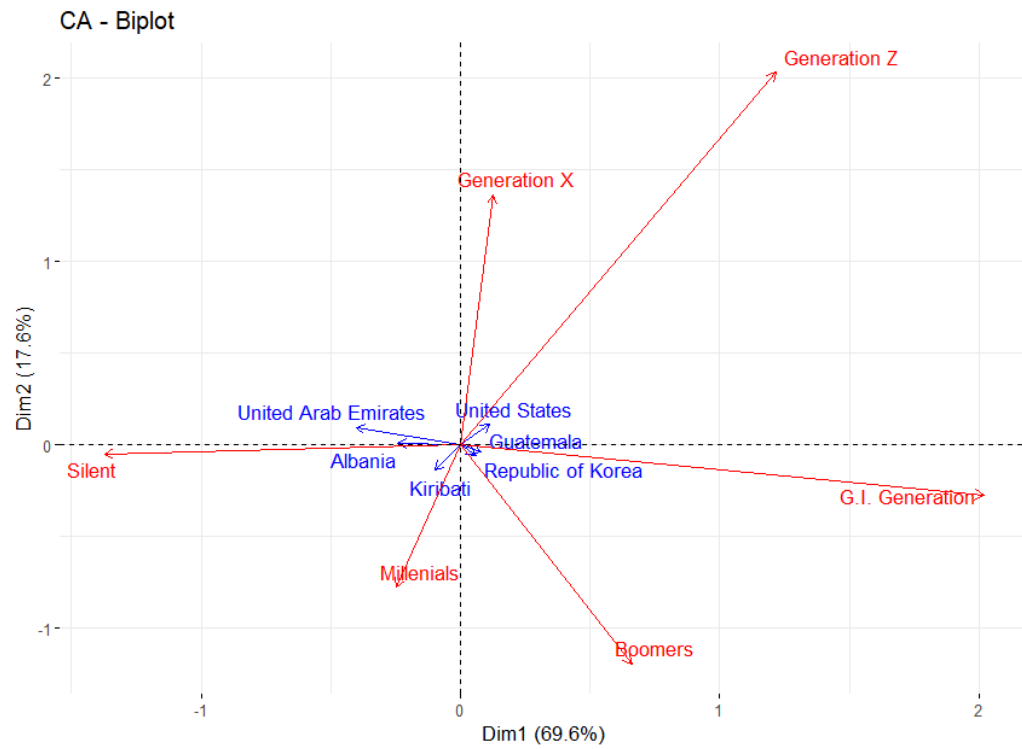
```
ca <- CA(table_generation_country, graph = TRUE)
```



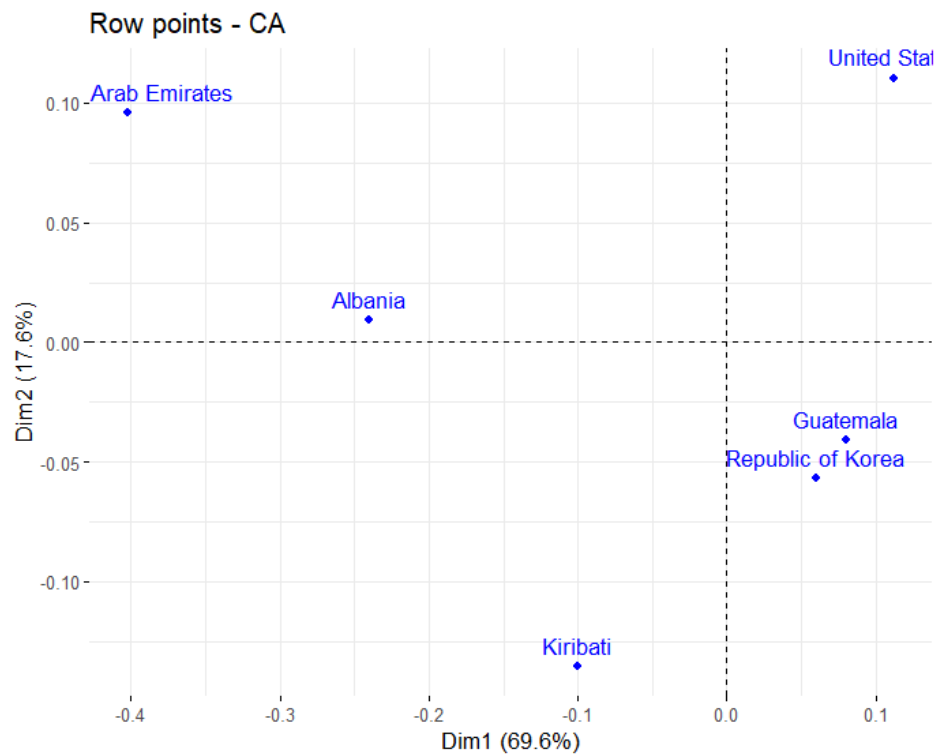
Interpretasi biplot variabel country dan generation

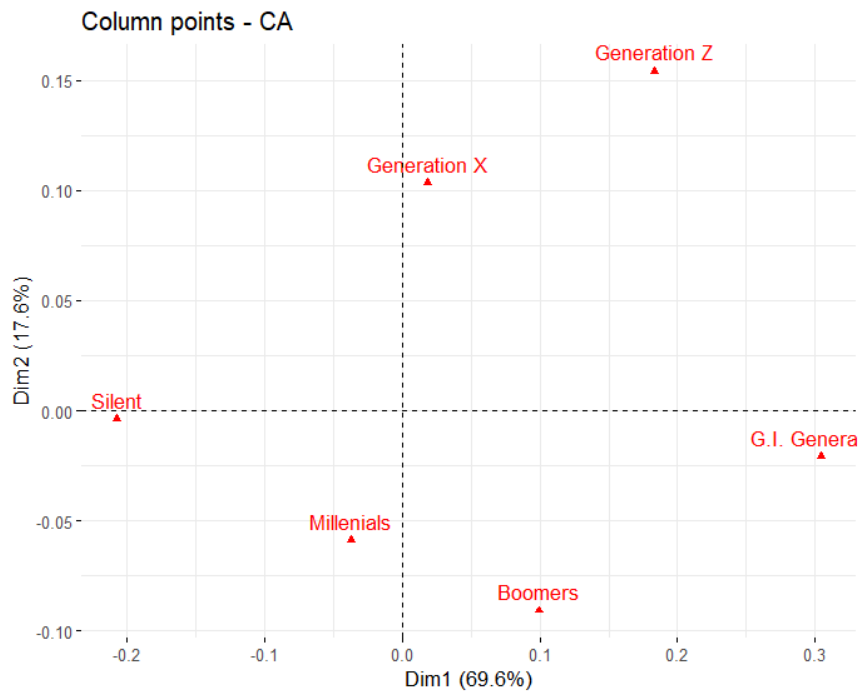
- Dimensi 1 (69.55%): Dimensi pertama menjelaskan 69.55% dari total variabilitas dalam data.
- Dimensi 2 (17.55%): Dimensi kedua menjelaskan 17.55% dari total variabilitas dalam data.
- Generasi Z dan X: Lebih tinggi kasus bunuh diri di Amerika Serikat.
- Silent Generation: Lebih tinggi kasus bunuh diri di Albania.
- Millennials: Lebih tinggi kasus bunuh diri di Korea Selatan.
- G.I. Generation dan Boomers: Lebih tinggi kasus bunuh diri di Guatemala.
- United States: Tinggi kasus bunuh diri di Generasi Z dan X.
- Republic of Korea: Tinggi kasus bunuh diri di Millennials.
- Guatemala: Tinggi kasus bunuh diri di G.I. Generation dan Boomers.
- Kiribati: Pola kasus bunuh diri berbeda atau lebih rendah secara keseluruhan.

Biplot asimetris



Visualisasi row profile dan column profiles secara terpisah





9. Plot kontribusi

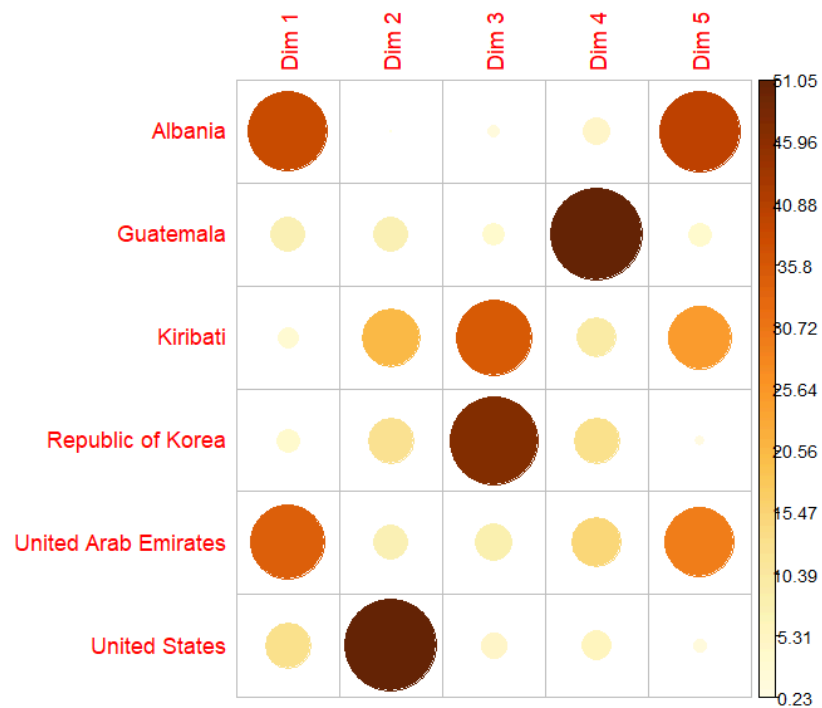
Baris

Kontribusi baris dalam persen

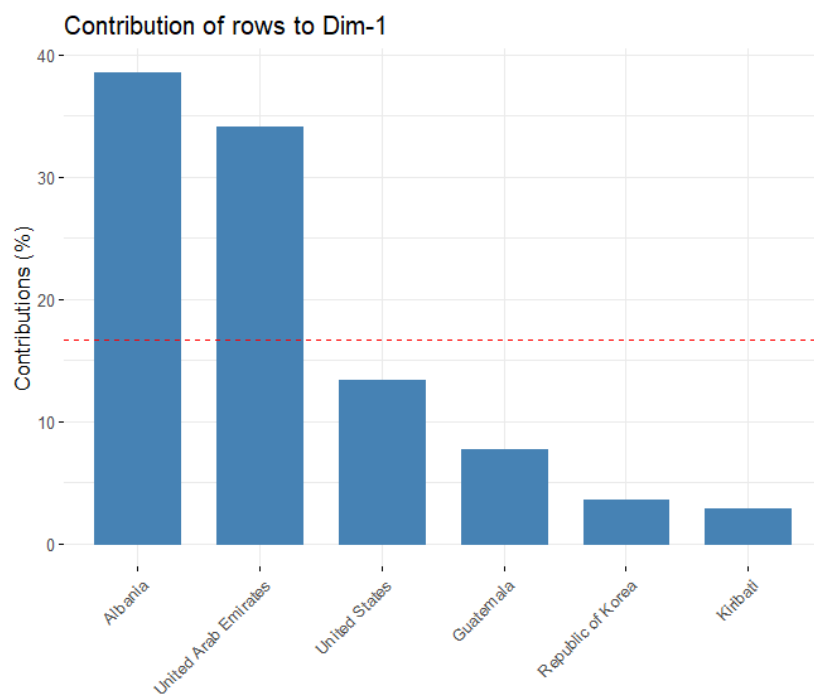
```
> head(row$contrib)
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Albania	38.566163	0.2281274	1.197095	4.719143	40.0894716
Guatemala	7.643635	7.8011305	3.280701	50.972940	3.5015945
Kiribati	2.806265	20.4741164	35.550632	10.013483	24.7555031
Republic of Korea	3.587636	12.7378573	46.976112	13.137348	0.7610469
United Arab Emirates	34.073351	7.7120430	8.499258	15.383252	29.5320950
United States	13.322950	51.0467254	4.496202	5.773834	1.3602889

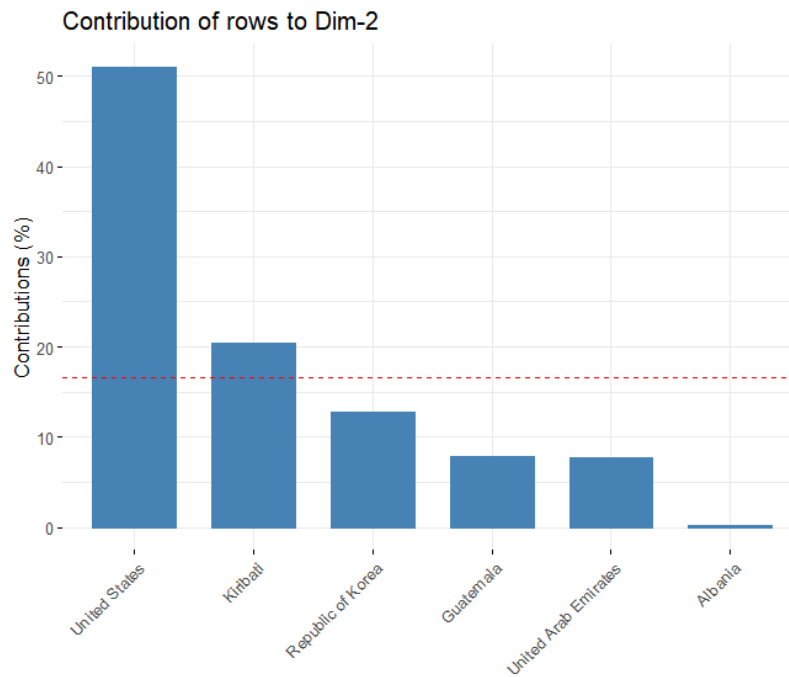
Berikut ini adalah korelasi plot untuk baris



Berikut ini adalah bar plot untuk melihat kontribusi baris dari dimensi 1 dan 2.



Terlihat Albania berkontribusi paling besar untuk dim 1 diikuti oleh Arab dan Amerika.



Untuk dim 2, Amerika memiliki kontribusi terbesar diikuti oleh Kiribati.

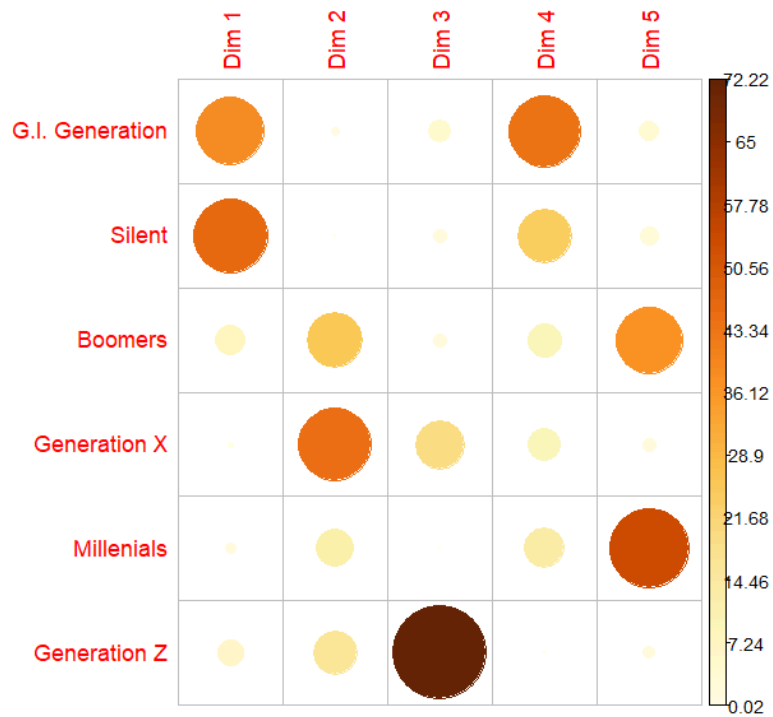
Kolom

Kontribusi kolom dalam persen

```
> head(col$contrib)
```

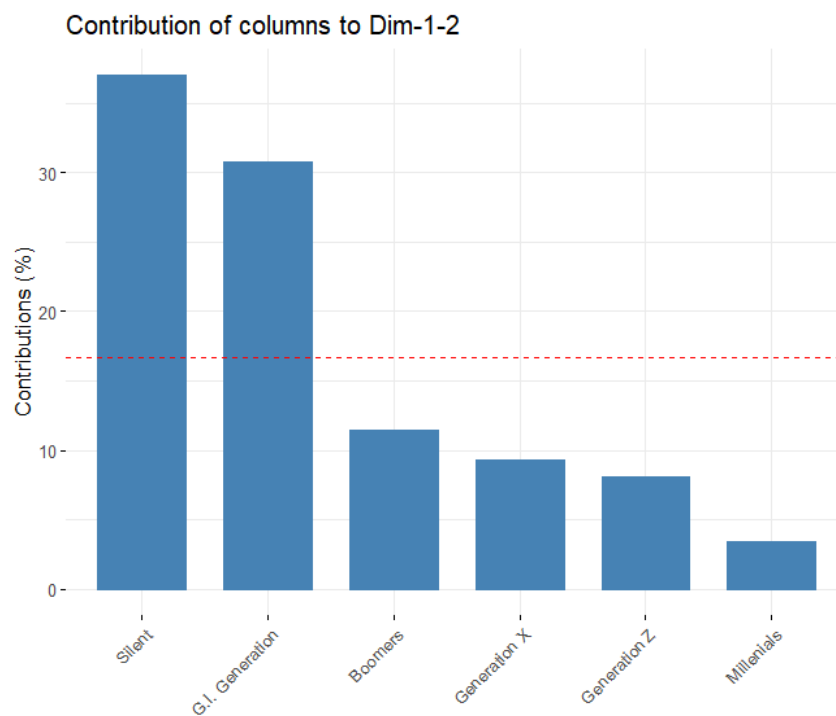
	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
G.I. Generation	38.2959686	0.71278141	4.49751364	43.48243442	3.611302
Silent	46.3598045	0.05845456	1.64866965	24.50578956	2.827282
Boomers	7.8348871	25.83914431	1.77509279	9.39509229	37.155783
Generation X	0.3716575	44.76669702	19.83364666	9.24827105	1.779728
Millenials	1.2160481	12.09716078	0.02435675	13.28862098	53.373813
Generation Z	5.9216343	16.52576191	72.22072051	0.07979171	1.252092

Berikut ini adalah corplot



- **GI Generation:** Berkontribusi signifikan di Dimensi 4.
- **Silent:** Kontribusi besar di Dimensi 1.
- **Boomers:** Kontribusi signifikan di Dimensi 5.
- **Gen X:** Kontribusi besar di Dimensi 2.
- **Millenials:** Kontribusi terbesar di Dimensi 5.
- **Gen Z :** Kontribusi signifikan di Dimensi 3.

Berikut ini adalah bar plot gabungan dim 1 dan 2 untuk kontribusi kolom



Terlihat bahwa generasi Silent memiliki kontribusi terbesar di kedua dimensi diikuti oleh GI Generation.

Bagian 5. Penutup

Dalam penelitian ini, penulis telah menerapkan analisis korespondensi sederhana untuk memahami hubungan antara perbedaan umur dan negara, umur dan generasi serta generasi dan negara terhadap tingkat bunuh diri. Dari hasil visualisasi data dan analisis data dapat disimpulkan bahwa di tiap perbedaan negara, generasi dan usia memiliki pola bunuh diri yang berbeda. Negara Amerika adalah negara dengan kasus bunuh diri terbanyak secara keseluruhan. Sementara Kiribati adalah negara dengan kasus bunuh diri paling sedikit. Untuk generasi, secara keseluruhan generasi Boomers (1946-1964) adalah generasi yang paling banyak melakukan bunuh diri. Sementara Gen Z (1997-2012) adalah generasi dengan kasus bunuh diri paling sedikit. Untuk kelompok usia, kasus terbanyak adalah 55-74 tahun dan paling sedikit adalah 5-14 tahun. Sedangkan untuk gender, pria memiliki kasus bunuh diri jauh lebih tinggi dibandingkan dengan wanita.

Hal tersebut dipengaruhi oleh banyak faktor. Tekanan hidup dan beban kerja yang meningkat seiring dengan bertambahnya usia, kesenjangan sosial yang tinggi di suatu negara, dan situasi dimana kondisi seseorang dilahirkan dapat menjadi faktor utama tindakan bunuh diri.

Penelitian ini telah memberikan wawasan berharga mengenai hubungan antara perbedaan generasi, usia, dan negara terhadap tingkat bunuh diri melalui analisis korespondensi sederhana. Temuan-temuan ini dapat digunakan sebagai dasar untuk mengembangkan strategi pencegahan bunuh diri yang lebih efektif dan terfokus, serta mendorong penelitian lebih lanjut untuk mengatasi masalah bunuh diri di berbagai kelompok usia dan negara.

Dengan memahami pola-pola ini, diharapkan kita dapat mengambil langkah-langkah konkret untuk mengurangi tingkat bunuh diri dan meningkatkan kesehatan mental global.

Bagian 6. Lampiran

Sumber data:

<https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016>

Link google colab:

https://colab.research.google.com/drive/188LXK_dHfXuQQOVpDwlkLQ6hOwGsOAxr?usp=sharing

Code R dan data bersih:

<https://drive.google.com/drive/folders/1OkHxIkwDGBITujkDhK-cf74ewjA0gmqN?usp=sharing>

Referensi:

Rencher & Christensen. 2012. Method of Multivariate Analysis. 3rd ed. Wiley & Sons

<http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/113-ca-correspondence-analysis-in-r-essentials#specification-in-ca>