

Analisis Emisi Karbon Dioksida

Geraldus Wilsen

Soal Unduhlah (download) 1 dataset (bebas) dari Kaggle (<https://www.kaggle.com/>) atau UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.php>) kemudian identifikasi jenis atributnya dan visualisasikan menggunakan teknik visualisasi yang sudah dijelaskan pada modul ini.

Jawaban Pada tugas kali ini, saya telah mendownload dataset “CO2 Emissions Around the World” melalui Kaggle (<https://www.kaggle.com/datasets/koustavghosh149/co2-emission-around-the-world/code?resource=download>). Secara singkat, melalui dataset ini kita bisa melihat bagaimana tingkat emisi karbon dari tahun 1990 - 2019 di seluruh dunia. Data ini sudah melalui **tahap pre - processing yang saya lakukan dengan menggunakan bantuan pyhton dan excel**. Secara detail dapat dilihat melalui tabel berikut:

```
setwd("C:/Users/LENOVO/OneDrive - Universitas Airlangga/SEMESTER 3/EVD")
data = read.csv("CO2_emission.csv")
head(data)
```

##	Country.Name	country_code	Region	X1990		
## 1	Afghanistan	AFG	South Asia	0.1917451		
## 2	Angola	AGO	Sub-Saharan Africa	0.5536620		
## 3	Albania	ALB	Europe & Central Asia	1.8195416		
## 4	Andorra	AND	Europe & Central Asia	7.5218317		
## 5	United Arab Emirates	ARE	Middle East & North Africa	30.1951886		
## 6	Argentina	ARG	Latin America & Caribbean	3.0755413		
##	X1991	X1992	X1993	X1994	X1995	X1996
## 1	0.1676816	0.09595774	0.08472111	0.07554583	0.06846796	0.06258803
## 2	0.5445386	0.54355722	0.70898423	0.83680440	0.91214149	1.07216847
## 3	1.2428102	0.68369983	0.63830704	0.64535519	0.60543625	0.61236736
## 4	7.2353792	6.96307870	6.72417752	6.54157891	6.73347949	6.99159455
## 5	31.7784962	29.08092584	29.27567777	30.84933296	31.12501806	30.92802588
## 6	3.2020313	3.21897372	3.24579528	3.25294529	3.24134189	3.47695321
##	X1997	X1998	X1999	X2000	X2001	X2002
## 1	0.05682662	0.05269086	0.04015697	0.0365737	0.03378536	0.04557366
## 2	1.08663697	1.09182531	1.10985966	0.9880774	0.94182891	0.89557767
## 3	0.46692147	0.57215370	0.95535931	1.0262131	1.05549588	1.23237878
## 4	7.30744115	7.63953851	7.92319165	7.9522863	7.72154906	7.56623988
## 5	30.48633262	29.66358052	28.88710798	27.0351591	29.43026994	28.50146173
## 6	3.53698995	3.67879380	3.68851827	3.5873920	3.36037118	3.11742483
##	X2003	X2004	X2005	X2006	X2007	X2008
## 1	0.05151838	0.04165539	0.06041878	0.06658329	0.06531235	0.1284166
## 2	0.92486944	0.93026295	0.81353929	0.82184008	0.81175351	0.8886580
## 3	1.33898498	1.40405869	1.33820940	1.33999574	1.39393137	1.3843112
## 4	7.24241557	7.34426233	7.35378001	6.79054277	6.53104692	6.4393039
## 5	27.96926982	27.03893822	25.38238104	22.93510429	21.37028576	22.0114692
## 6	3.35172371	3.67297400	3.75363924	3.94249129	4.10262964	4.1723885

```
##      X2009      X2010      X2011      X2012      X2013      X2014      X2015
## 1  0.1718624  0.2436140  0.2965062  0.2592953  0.1856237  0.1462356  0.1728967
## 2  0.9394040  0.9761842  0.9855223  0.9506959  1.0362939  1.0997791  1.1350441
## 3  1.4414936  1.5276237  1.6694232  1.5032405  1.5336300  1.6683374  1.6037751
## 4  6.1566875  6.1571978  5.8508861  5.9446542  5.9428004  5.8071277  6.0261818
## 5 19.8323489 19.0397698 18.5094574 19.2078011 20.0556476 20.0516980 21.0776420
## 6  3.8675700  4.0996897  4.2809894  4.2642238  4.3421151  4.2090956  4.3019139
##      X2016      X2017      X2018      X2019      Class
## 1  0.1497893  0.1316946  0.1632953  0.1598244 Below Average
## 2  1.0318113  0.8133007  0.7776749  0.7921371 Below Average
## 3  1.5576644  1.7887861  1.7827389  1.6922483 Below Average
## 4  6.0806003  6.1041339  6.3629754  6.4812174 Above Average
## 5 21.4806686 20.7690223 18.3906781 19.3295633 Above Average
## 6  4.2018458  4.0713084  3.9757720  3.7406503 Below Average
```

```
ls(data)
```

Identifikasi Atribut

```
## [1] "Class"          "Country.Name"  "country_code" "Region"        "X1990"
## [6] "X1991"          "X1992"         "X1993"         "X1994"         "X1995"
## [11] "X1996"          "X1997"         "X1998"         "X1999"         "X2000"
## [16] "X2001"          "X2002"         "X2003"         "X2004"         "X2005"
## [21] "X2006"          "X2007"         "X2008"         "X2009"         "X2010"
## [26] "X2011"          "X2012"         "X2013"         "X2014"         "X2015"
## [31] "X2016"          "X2017"         "X2018"         "X2019"
```

Melalui tabel diatas, saya telah mengidentifikasi setiap atribut variabel sebagai berikut:

Nama Variabel	Jenis Atribut	Alasan
Country Name	Kualitatif Nominal	Variabel "Country Name" menunjukkan nama - nama setiap negara yang ada di dunia, artinya variabel ini tidak mengukur besaran tetapi hanya sebagai penggolongan saja
Country Code	Kualitatif Nominal	Variabel "Country Code" menunjukkan kode setiap negara yang ada di dunia, artinya variabel ini tidak mengukur besaran tetapi hanya sebagai penggolongan saja

Nama Variabel	Jenis Atribut	Alasan
Region	Kualitatif Nominal	Variabel “Region” menunjukkan klasifikasi area setiap negara yang ada di dunia, artinya variabel ini tidak mengukur besaran tetapi hanya sebagai penggolongan saja
Class	Kualitatif Biner	Variabel “Class” menunjukkan apakah kadar emisi suatu karbon suatu negara diatas atau dibawah rata - data, sehingga tidak bisa dilakukan perhitungan serta memenuhi kriteria data biner
X1990 - X2019	Kuantitatif Kontinu (Skala: Rasio)	Variabel “1990” sampai dengan variabel “2019” menunjukkan data numerik (tingkat CO2) setiap negara sepanjang tahun 1990 sampai 2019. Tingkat CO 2 memiliki nilai 0 mutlak. Artinya ketika dikonversikan ke satuan ukur lain atau diukur dengan alat ukur lain, ketika hasilnya 0 maka akan selalu tetap 0. Maka dari itu, variabel ini dikategorikan dalam skala perhitungan rasio, dan juga bersifat kontinu

Visualisasi Pertama, saya telah menginstall beberpaa library yang akan saya gunakan sebagai berikut:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.1
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.2.1
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.1
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v tibble  3.1.8      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## v purrr   0.3.4

## Warning: package 'tibble' was built under R version 4.2.1

## Warning: package 'tidyr' was built under R version 4.2.1

## Warning: package 'readr' was built under R version 4.2.1

## Warning: package 'purrr' was built under R version 4.2.1

## Warning: package 'dplyr' was built under R version 4.2.1

## Warning: package 'stringr' was built under R version 4.2.1

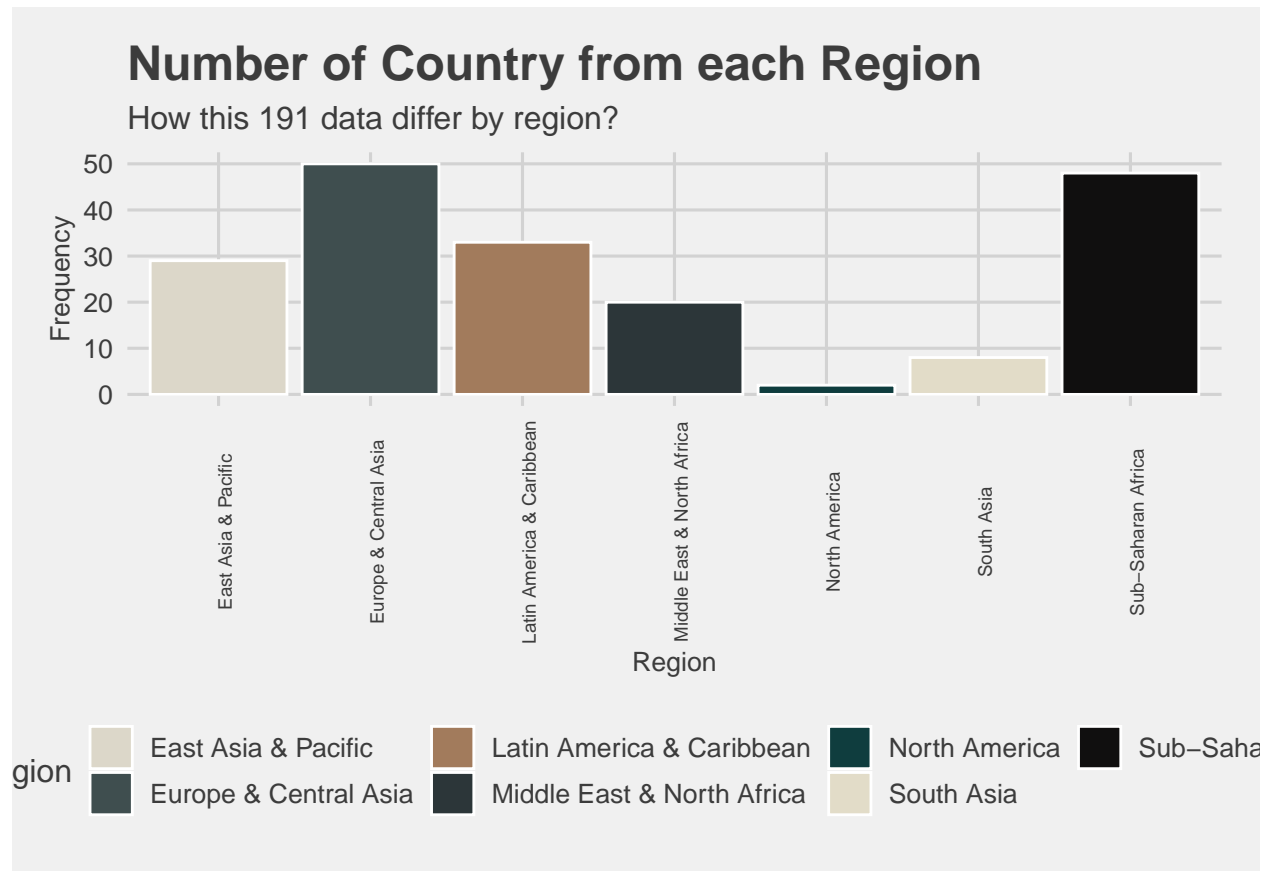
## Warning: package 'forcats' was built under R version 4.2.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

Setelah itu, saya menjawab beberapa pertanyaan berikut melalui visualisasi data:

1. Dari 191 negara di dunia, berapa jumlahnya jika dibagi berdasarkan region? (Bar Chart)

```
ggplot(data,aes(x=Region,fill = Region)) +
  geom_bar(color = "white")+
  labs(title = "Number of Country from each Region",
        subtitle = "How this 191 data differ by region?",
        x = "Region",
        y = "Frequency")+
  theme_fivethirtyeight()+
  theme(axis.text.x = element_text(angle=90,size = 7))+
  theme(axis.title = element_text(size = 10))+
  scale_fill_manual(values = c("East Asia & Pacific" = "#DCD7C9",
                                "Europe & Central Asia"= "#3F4E4F",
                                "Latin America & Caribbean" ="#A27B5C",
                                "Middle East & North Africa" = "#2C3639",
                                "North America" = "#0F3D3E",
                                "South Asia" = "#E2DCC8",
                                "Sub-Saharan Africa" = "#100F0F"))
```



Dari visualisasi diatas, saya mendapatkan kesimpulan dimana dari 191 data negara yang ada dalam dataset ini, region dengan negara paling banyak adalah “Europe & Central Asia” dengan 50 negara

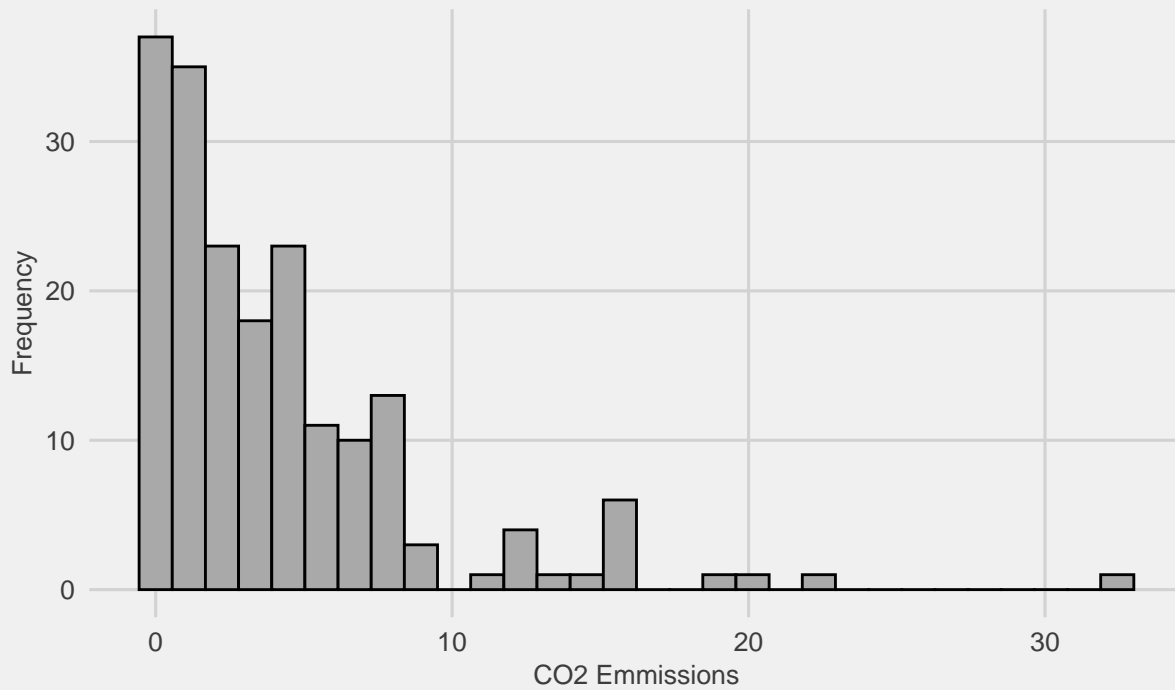
2. Kadar CO2 pada tahun 2019? (Histogram, Kernel Density Plot)

```
ggplot(data,aes(X2019)) +
  geom_histogram(fill = "dark gray",color= "black")+
  labs(x = "CO2 Emmissions",y = "Frequency",title = "CO2 Emmissions in 2019 Around the World", subtitle
  theme_fivethirtyeight()+
  theme(axis.title = element_text(size= 10))
```

‘stat_bin()’ using ‘bins = 30’. Pick better value with ‘binwidth’.

CO2 Emmissions in 2019 Around the World

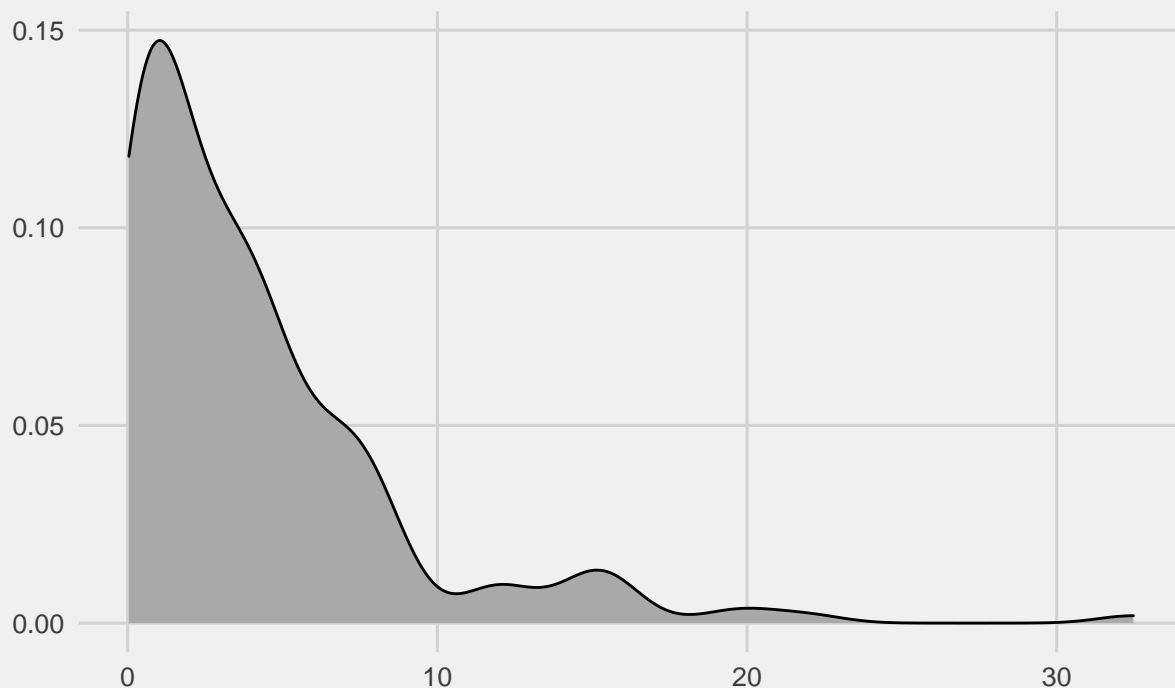
How is the distribution in 2019?



```
ggplot(data,aes(x = X2019))+  
  geom_density(fill = "dark gray")+  
  labs(x = "CO2 Emmissions",y = "Frequency",title = "CO2 Emmissions in 2019 Around the World", subtitle  
  theme_fivethirtyeight()
```

CO2 Emmissions in 2019 Around the World

How is the distribution in 2019?



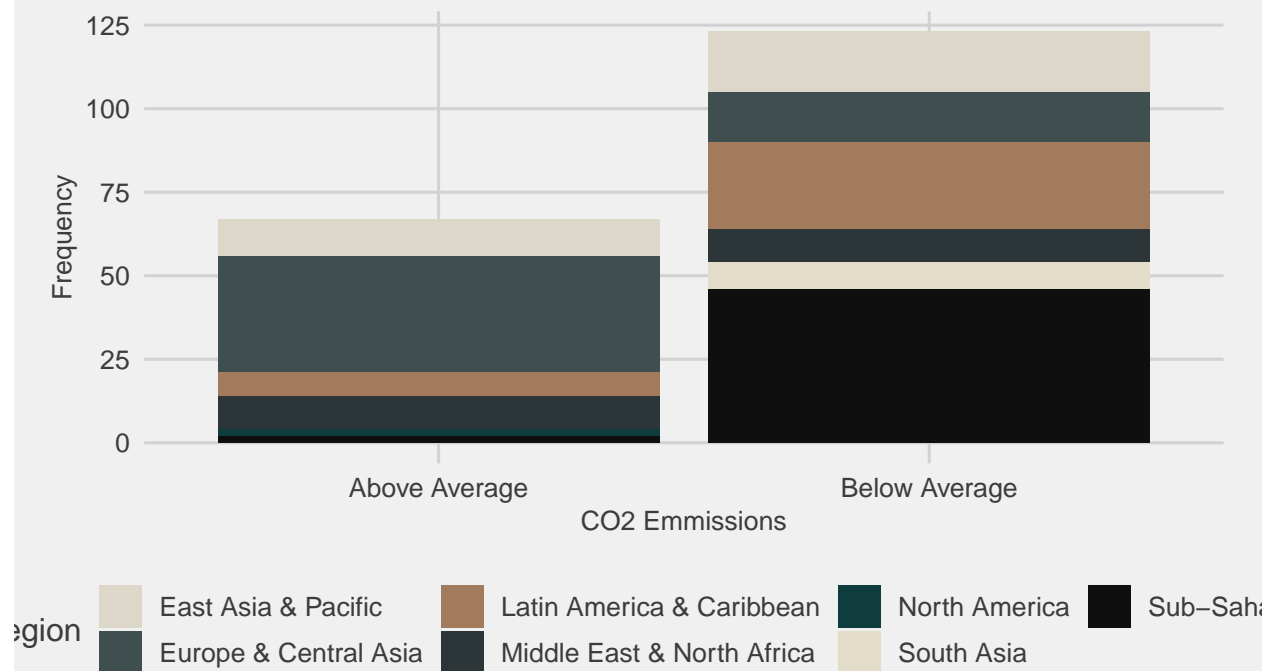
Melalui Histogram dan Kernel plot diatas, saya dapat menarik sebuah kesimpulan bahwa kadar CO2 dalam rentang 0 - 2 adalah yang terbanyak, dan grafik semakin landai, artinya semakin sedikit jumlah negara yang mengeluarkan emisi karbondioksida dalam kadar tinggi

3. Pada tahun 2019, berapa banyak negara yang masuk dalam kategori kadar CO2 dibawah maupun diatas rata - rata (4.134273353 metric tons per capita)? Dari region mana yang paling besar? (Stacked, Grouped, Segmented bar charts)

```
ggplot(data,aes(Class,fill = Region)) +  
  geom_bar(position= "stack")+  
  labs(x = "CO2 Emmissions",y = "Frequency",title = "CO2 Emmissions in 2019 Around the World", subtitle = "How is the distribution in 2019?")  
  theme_fivethirtyeight()+  
  theme(axis.title = element_text(size= 10))+  
  scale_fill_manual(values = c("East Asia & Pacific" = "#DCD7C9",  
                                "Europe & Central Asia"= "#3F4E4F",  
                                "Latin America & Caribbean" = "#A27B5C",  
                                "Middle East & North Africa" = "#2C3639",  
                                "North America" = "#0F3D3E",  
                                "South Asia" = "#E2DCC8",  
                                "Sub-Saharan Africa" = "#100F0F")))
```

CO2 Emmissions in 2019 Around the World

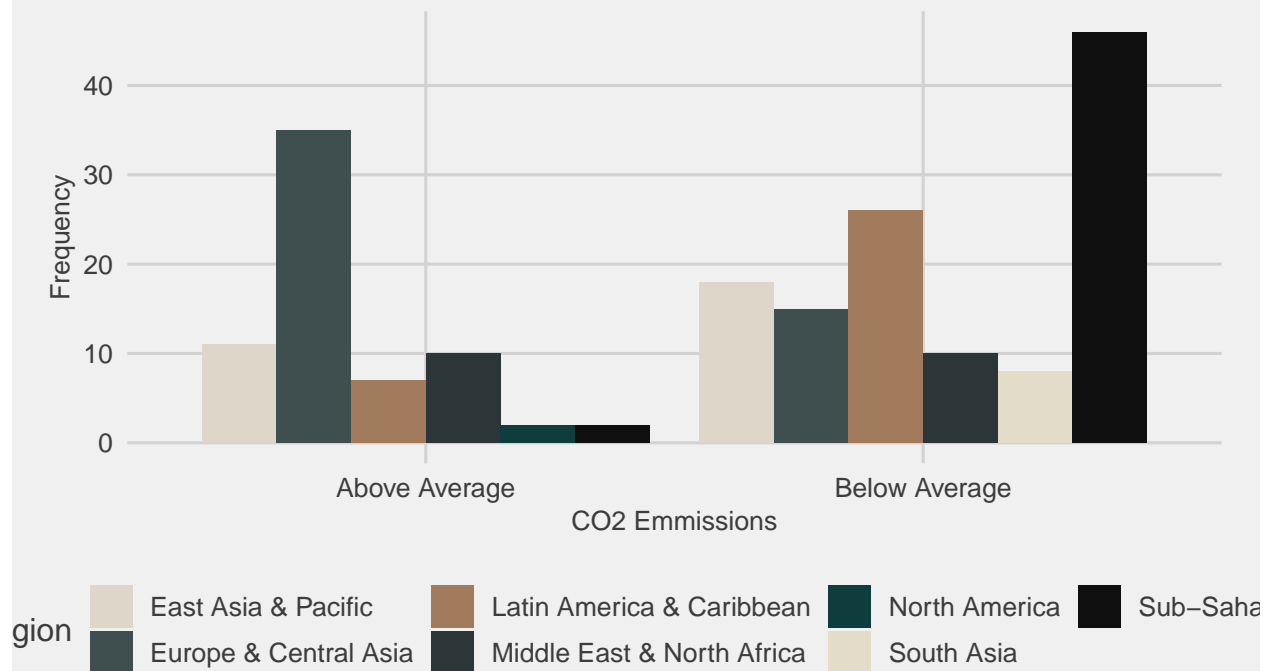
How is the distribution in 2019?



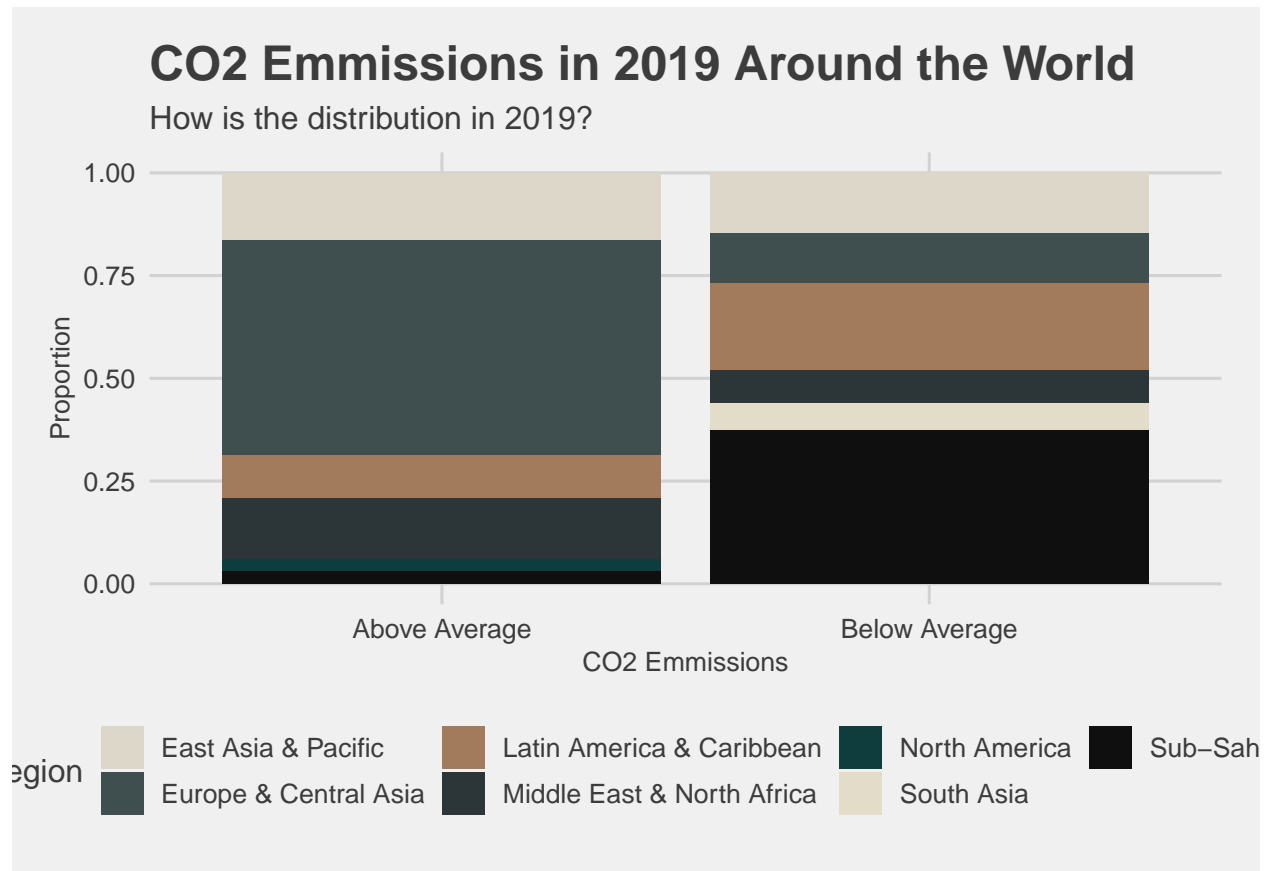
```
ggplot(data,aes(Class,fill = Region)) +
  geom_bar(position= "dodge")+
  labs(x = "CO2 Emmissions",y = "Frequency",title = "CO2 Emmissions in 2019 Around the World", subtitle = "How is the distribution in 2019?")
  theme_fivethirtyeight()+
  theme(axis.title = element_text(size= 10))+
  scale_fill_manual(values = c("East Asia & Pacific" = "#DCD7C9",
                                "Europe & Central Asia"= "#3F4E4F",
                                "Latin America & Caribbean" = "#A27B5C",
                                "Middle East & North Africa" = "#2C3639",
                                "North America" = "#0F3D3E",
                                "South Asia" = "#E2DCC8",
                                "Sub-Saharan Africa" = "#100F0F"))
```


CO2 Emmissions in 2019 Around the World

How is the distribution in 2019?



```
ggplot(data,aes(Class,fill = Region)) +
  geom_bar(position= "fill")+
  labs(x = "CO2 Emmissions",y = "Proportion",title = "CO2 Emmissions in 2019 Around the World", subtitle = "How is the distribution in 2019?")
  theme_fivethirtyeight()+
  theme(axis.title = element_text(size= 10))+
  scale_fill_manual(values = c("East Asia & Pacific" = "#DCD7C9",
                                "Europe & Central Asia"= "#3F4E4F",
                                "Latin America & Caribbean" = "#A27B5C",
                                "Middle East & North Africa" = "#2C3639",
                                "North America" = "#0F3D3E",
                                "South Asia" = "#E2DCC8",
                                "Sub-Saharan Africa" = "#100F0F"))
```



Melalui Stacked, Grouped dan Segmented Bar Chart diatas, saya dapat menarik sebuah kesimpulan dimana negara - negara yang memiliki kadar emisi karbon dioksida diatas rata” terbanyak dari region Europe & Central Asia, sedangkan Sub - Saharan sebagai region dengan jumlah negara paling banyak penyumbang emisi karbon dibawah rata- rata.

4. Bagaimana tren kadar CO2 Indonesia dibandingkan negara ASEAN lainnya?(Multiple Line Chart)

```
trend = read.csv("pivotdata.csv")
head(trend)
```

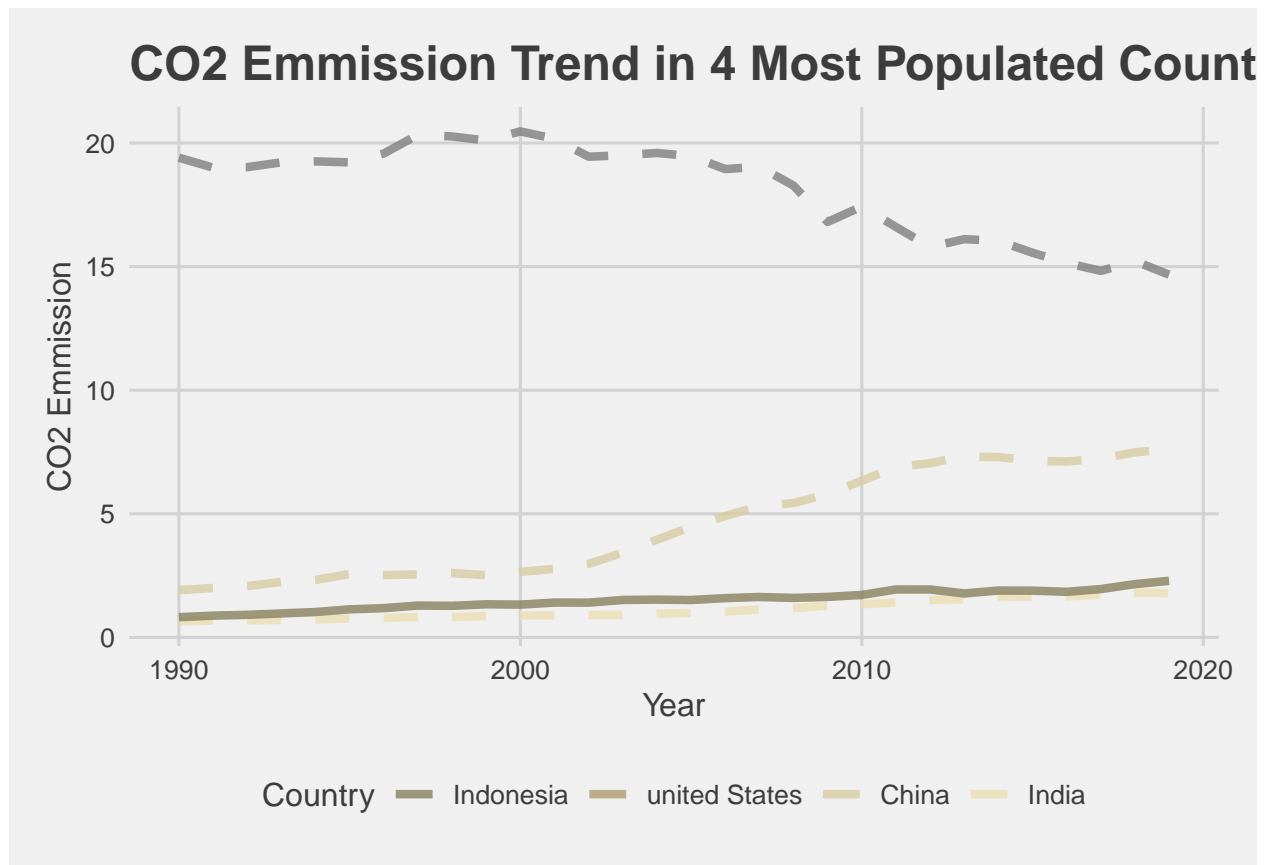
```
##      Country Year      CO2
## 1      China 1990 1.9145426
## 2      India 1990 0.6453616
## 3  Indonesia 1990 0.8187378
## 4 United States 1990 19.4073463
## 5      China 1991 2.0005388
## 6      India 1991 0.6813062
```

```
trend %>%
  mutate(isIndonesia = (Country == "Indonesia")) %>%
  ggplot(aes(x = Year, y = CO2, color = Country)) +
  geom_line(aes(linetype = isIndonesia), size = 1.5, alpha = 0.8) +
  labs(title = "CO2 Emmission Trend in 4 Most Populated Countries",
       x = "Year",
       y = "CO2 Emmission",
```

```

color = "Country")+
theme_fivethirtyeight()+
theme(axis.title = element_text())+
scale_linetype_manual(values=c("dashed","solid"),guide="none")+
scale_color_manual(values = c("Indonesia" = "#87805E",
                              "united States"= "#B09B71",
                              "China" = "#D8CCA3",
                              "India" = "#EDDFB3"))

```



Melalui visualisasi ini, saya dapat menjawab hipotesis bahwa jumlah masyarakat sebuah negara memengaruhi jumlah emisi karbon yang dihasilkan. Walaupun tidak 100% dapat dikatakan memengaruhi, namun saya mampu menyimpulkan bahwa majunya industrialisasi sebuah negara memengaruhi jumlah emisi karbon dioksida. Hal ini ditunjukkan oleh Amerika pada tahun 1990 merupakan negara paling maju dalam industri, sehingga memiliki emisi karbon sebesar kurang lebih 20 million per capita. Disisi lain, pada saat itu, China, Indonesia dan India masih belum semasif Amerika. Namun seiring perkembangan zaman. Ketiga negara tersebut mulai berkembang dan terlihat sangat jelas China sebagai negara yang saat ini maju dalam industri, meningkat secara cukup signifikan pada tahun 2019. Begitu pula dengan India dan Indonesia yang kian meningkat setiap tahunnya.

5. Bagaimana statistik numerik antara region dengan rata - rata kadar CO2 dalam 29 tahun ke belakang? (Ridgeline Plot)

```

avg = read.csv("avg.csv")
head(avg)

```

##	Country.Name	country_code	Region	Average
----	--------------	--------------	--------	---------

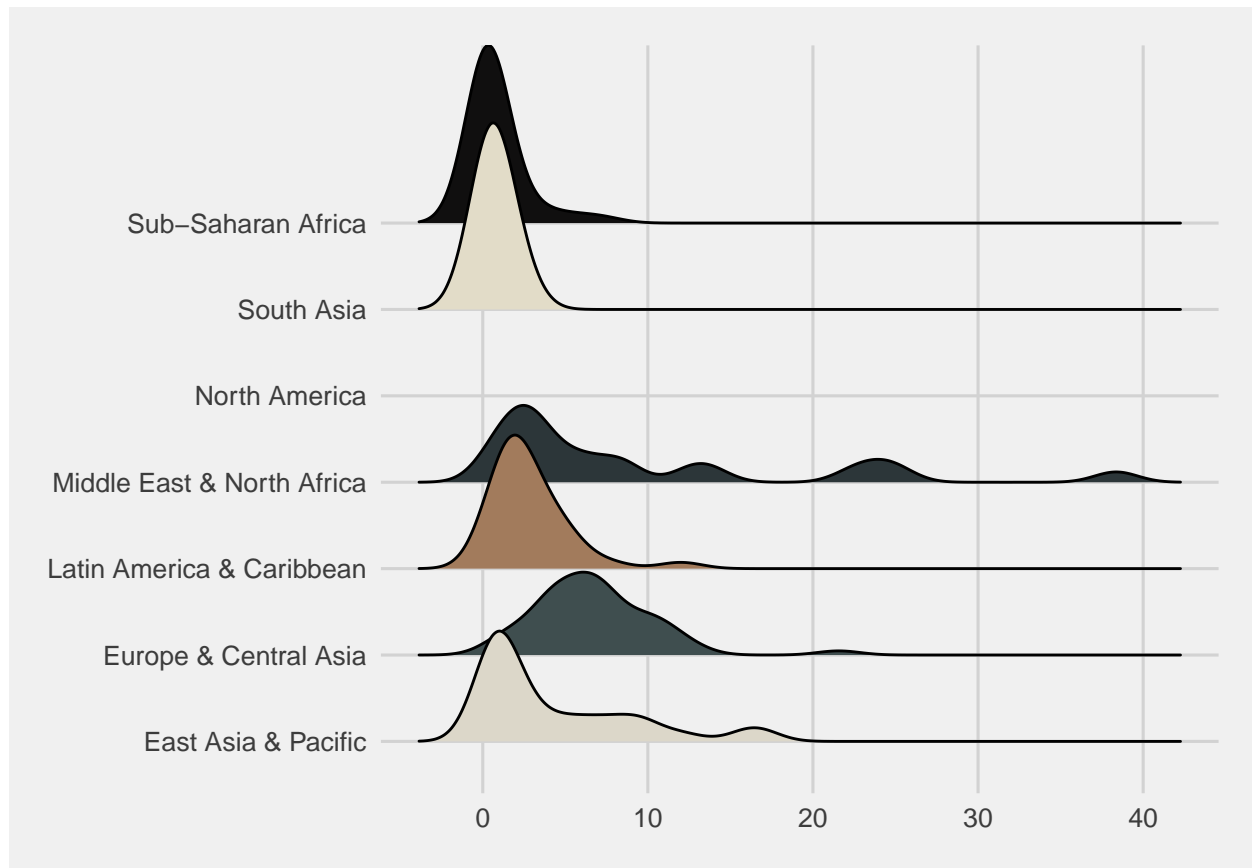
```
## 1      Australia      AUS East Asia & Pacific 16.827027
## 2    Brunei Darussalam BRN East Asia & Pacific 16.061935
## 3      China         CHN East Asia & Pacific  4.550064
## 4      Fiji         FJI East Asia & Pacific  1.278046
## 5 Micronesia, Fed. Sts. FSM East Asia & Pacific  1.291522
## 6      Indonesia    IDN East Asia & Pacific  1.513296
```

```
library(ggbridges)
```

```
## Warning: package 'ggbridges' was built under R version 4.2.1
```

```
ggplot(avg,aes(x = Average,
               y = Region,
               fill = Region))+
  geom_density_ridges()+
  theme_ridges()+
  labs("CO2 Emmissions")+
  theme_fivethirtyeight()+
  theme(legend.position = "none")+
  scale_fill_manual(values = c("East Asia & Pacific" = "#DCD7C9",
                              "Europe & Central Asia"= "#3F4E4F",
                              "Latin America & Caribbean" = "#A27B5C",
                              "Middle East & North Africa" = "#2C3639",
                              "North America" = "#0F3D3E",
                              "South Asia" = "#E2DCC8",
                              "Sub-Saharan Africa" = "#100F0F"))
```

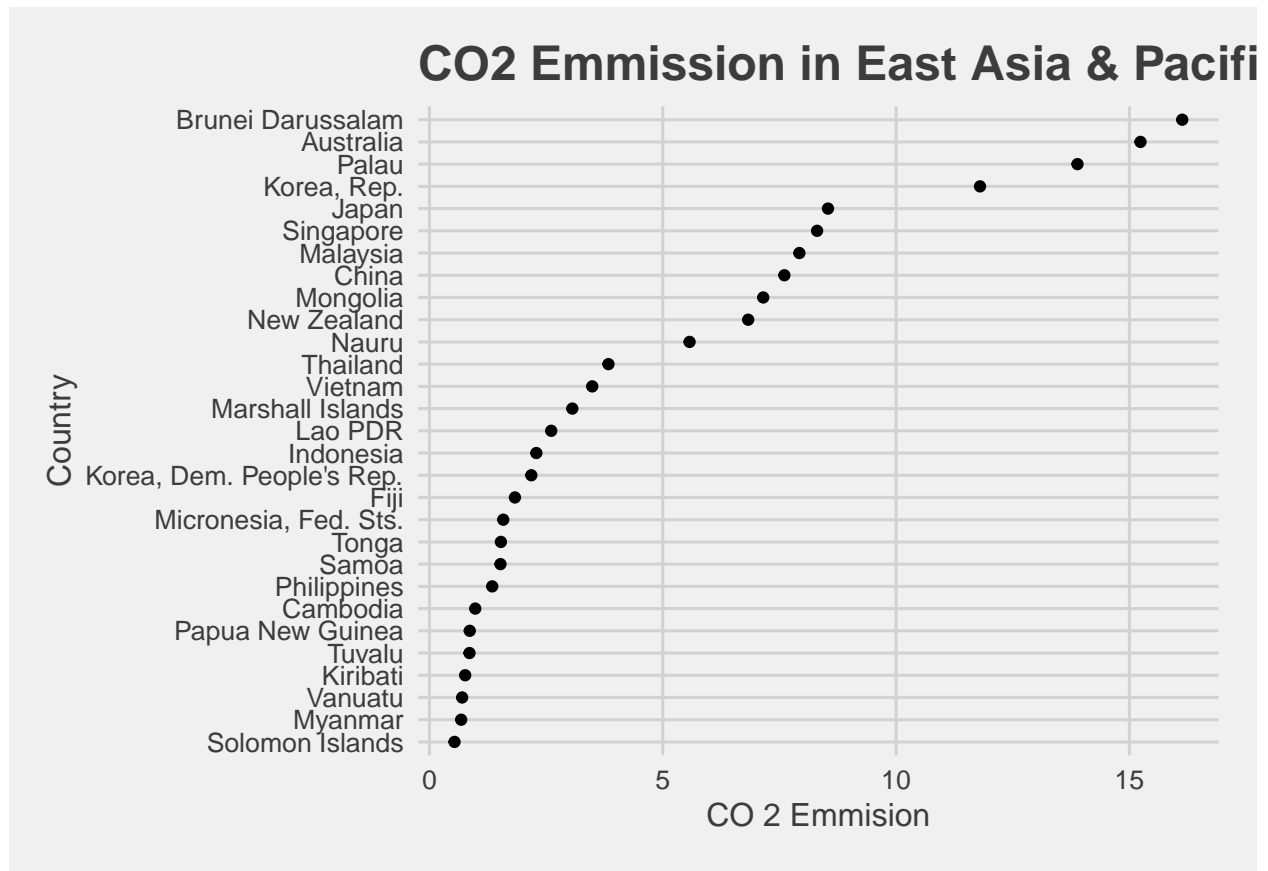
```
## Picking joint bandwidth of 1.3
```



Melalui ridgeline plot diatas, saya mendapatkan beberapa kesimpulan, dimana Sub Saharan Africa, Latin America, serta South Asia memiliki distribusi emisi karbon dioksida dalam rentang rendah. Sedangkan terdapat beberapa negara yang menghasilkan emisi karbon cukup tinggi seperti di region Middle East & North Africa. Walaupun begitu, jika dibandingkan dengan region Europe & Central Asia, negara - negara di kawasan ini memiliki rata” yang jauh lebih tinggi dibandingkan kawasan lainnya. Terakhir ada East Asia & Pacific, yang mayoritas emisi karbon nya dalam rentang rendah, namun adapula beberapa negara yang menghasilkan emisi cukup tinggi di kisaran 10 - 20.

6. Bagaimana statistik numerik kadar CO2 di seluruh region East Asia & Pacific pada tahun 2019? (Cleveland Dot Chart)

```
data1 = subset(data,Region == "East Asia & Pacific")
ggplot(data1,
  aes(x=X2019,y = reorder(Country.Name,X2019)))+
  geom_point()+
  labs(title = "CO2 Emmission in East Asia & Pacific Region",
    x = "CO 2 Emmision",
    y = "Country") +
  theme_fivethirtyeight()+
  theme(axis.title = element_text())
```



Dari visualisasi ini, saya mendapatkan kesimpulan bahwa pada tahun 2019 mayoritas negara - negara di East Asia & Pacific menghasilkan emisi karbon dengan kadar 0 - 5. Adapun negara sebagai penyumbang emisi karbon dengan kadar paling rendah adalah Kepulauan Solomon, dan yang paling tinggi adalah Brunei Darussalam