

# Report per l'esame di Laboratorio di Analisi dei Dati (M-Z)

Giulio Mazzarelli

## **Nomi dei dataset utilizzati:**

*Country\_region*

*world-production-statistics*

*world-refined-production-statistics*

*coal-production-data*

*world-natural-gas-production-statistics*

*world-electricity-production-statistics*

*renewable-in-electricity-production-share*

*wind-solar-share-electricity-production*

*CO2-emissions-data-from-fuel-combustion*

## **Nome degli script utilizzati:**

script\_finale.R

## **Fonte dei dati:**

<https://www.gapminder.org/data/geo/>

<https://yearbook.enerdata.net/>

selezionando i singoli menù (crude oil, oil products, natural gas, coal lignite, electricity, renewables, CO2 fuel combustion) è possibile scaricare i dataset in formato .xls

(per scaricare i dataset da yearbook.enerdata.net è necessario registrarsi al sito)

## Descrizione del fenomeno analizzato:

L'analisi in questione ha come obiettivo quello di analizzare i tipi di energia che vengono prodotti in ciascun Paese e in che porzione l'energia rinnovabile è, in tempi recenti (2018), sostenuta come fonte. È inoltre proposta la misura delle Mt di CO<sub>2</sub> come proxy del livello di salute ambientale di ciascun Paese nei termini che riguardano l'inquinamento atmosferico.

Si è scelto la quantità di CO<sub>2</sub> nell'aria perché rappresenta sia l'effetto primario dell'utilizzo di combustibili fossili sia la causa del problema "effetto serra".

Fonte: [https://www.cicap.org/n/articolo.php?id=274240#ref\\_17](https://www.cicap.org/n/articolo.php?id=274240#ref_17)

## Codebook:

Nome Variabile	Descrizione Variabile	Valori	Tipo
<b>country</b>	<i>Nome del Paese</i>	<i>Nome del Paese</i>	<i>Qualitativo sconnesso</i>
<b>region</b>	<i>Macroarea geografica del Paese</i>	<i>Africa, America, Asia, Europa</i>	<i>Qualitativo categorica</i>
<b>p.greggio</b>	<i>Produzione di greggio (Mt)</i>	<i>Numerica: 0 valore minimo, -</i>	<i>Quantitativa continua</i>
<b>p.raffinati</b>	<i>Produzione di raffinati (Mt)</i>	<i>Numerica: 0 valore minimo, -</i>	<i>Quantitativa continua</i>
<b>p.coalignite</b>	<i>Produzione di carbone e lignite (Mt)</i>	<i>Numerica: 0 valore minimo, -</i>	<i>Quantitativa continua</i>
<b>p.gas</b>	<i>Produzione di gas naturali (bcm)</i>	<i>Numerica: 0 valore minimo, -</i>	<i>Quantitativa continua</i>
<b>p.ele</b>	<i>Produzione di energia elettrica (TWh)</i>	<i>Numerica: 0 valore minimo, -</i>	<i>Quantitativa continua</i>
<b>s.eleinnov</b>	<i>Percentuale di energia rinnovabile sul totale di energia elettrica (%)</i>	<i>Numerica: 0 valore minimo, 100 valore massimo</i>	<i>Quantitativa continua</i>
<b>s.windsolr</b>	<i>Percentuale di energia eolica e solare sul totale di energia elettrica prodotta (%)</i>	<i>Numerica: 0 valore minimo, 100 valore massimo</i>	<i>Quantitativa continua</i>
<b>co2</b>	<i>Emissioni di CO derivanti dalla combustione di carburanti (Mt)</i>	<i>Numerica: 0 valore minimo, -</i>	<i>Quantitativa continua</i>

*Mt: Megatonnelate | 1,0e+09 Chilogrammi*

*bcm: billion cubic metres (miliardo di metri cubi)*

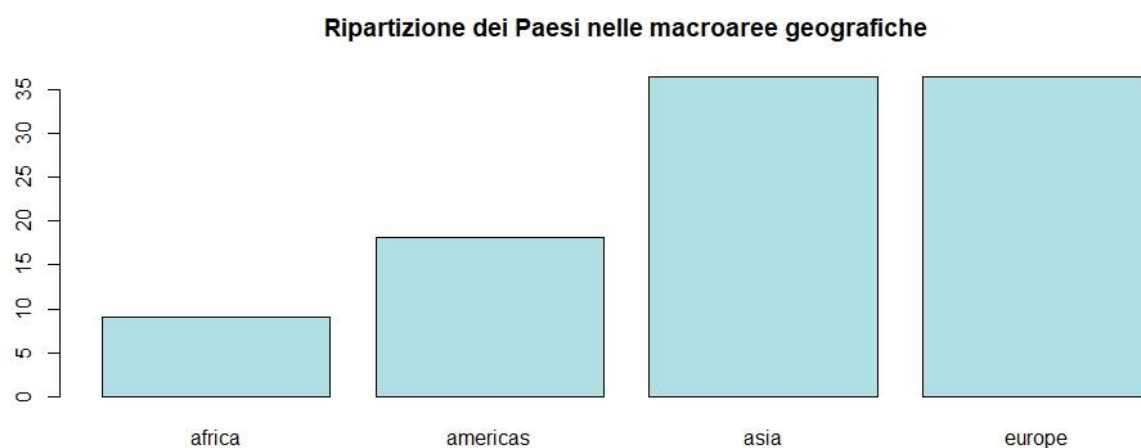
*TWh: Terawattora | 1,0e+12 Wattora*

## Azioni svolte in Excel:

1. Ho salvato tutti i dataset in un unico file Excel;
  2. Ho ripulito i dataset di tutti gli anni tranne che la colonna relativa all'anno 2018;
  3. Ho usato la funzione CERCA.VERT nel foglio "unionedati" per unire le diverse variabili che ho deciso di considerare;
  4. Ho modificato font, stile del carattere e colore delle celle relative al nome delle variabili e dei Paesi nel foglio "unionedati";
  5. Ho eliminato le righe relative ad aggregazioni di Paesi e continenti (World, G7, North America, Asia ecc.);
  6. Ho utilizzato la funzione CERCA.VERT per la variabile "region" che assegna ad ogni Paese la corrispettiva macroarea geografica;
  7. Non essendovi valori mancanti non è stato necessario applicare un filtro;
  8. Ho convertito i valori numerici delle colonne s.elerinnov e s.windsolar in valori percentuali (i valori non erano inizialmente in formato decimale, ho quindi diviso tutto per cento prima di utilizzare l'opzione fornita da Excel);
  9. Ho eseguito un copia e incolla trasposto della riga delle variabili dal foglio "unionedati" alla prima colonna del foglio "codebook";
  10. Ho fornito una breve descrizione e indicato il tipo per ciascuna variabile considerata;
  11. Da R ho esportato diverse tabelle, contenenti misure di sintesi, confronti tra variabili categoriche, correlazioni tra variabili e altri tipi di analisi;
  12. Una volta esportate le tabelle come file .txt le ho importate in Excel;
  13. Su Excel ho quindi apportato alcune modifiche estetiche e altre funzionali, come trasporre le tabelle in verticale o sostituire con l'apposito strumento "Trova e Sostituisci" i punti (che R utilizza come separatore dei decimali) in virgole. In tal modo ho potuto eseguire operazioni come la riduzione di cifre decimali mostrate o la conversione delle frequenze relative in valori percentuali;
  14. Ho quindi copiato e incollato le tabelle da Excel nel report.
-

## Analisi marginali:

### Com'è composto il dataset?



Le osservazioni, ovvero i Paesi, sono ripartiti nelle quattro macroaree geografiche nelle percentuali indicate dal grafico.

### Dati generali sulle variabili

	p.greggio	p.raffinati	p.coalignite	p.gas	p.ele	s.elerinnov	s.windsolar	co2
Min.	0	2.12	0	0	33.31	0	0	31.12
1st Qu.	1.44	21.15	0.02	2.44	97.24	0.12	0	100.53
Median	22.51	34.9	2.77	26.47	197.78	0.19	0.05	261.61
Mean	86.48	88.71	167.03	76.84	551.79	0.28	0.07	686.28
3rd Qu.	94.2	75.24	84.41	63.12	338.3	0.39	0.11	453
Max.	675.98	895.91	3474.49	864.07	7091.79	0.98	0.25	9466.5

#### Varianza

p.greggio	31372.56
p.raffinati	33054.59
p.coalignite	315583.03
p.gas	33832.2
p.ele	1772557.68
co2	2934359.61

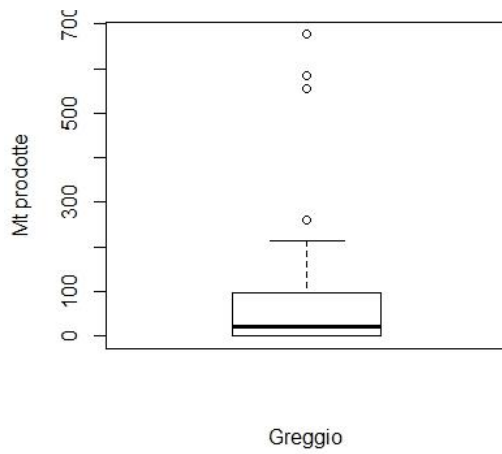
#### Deviazione Standard

p.greggio	177.12
p.raffinati	181.81
p.coalignite	561.77
p.gas	183.94
p.ele	1331.37
co2	1713

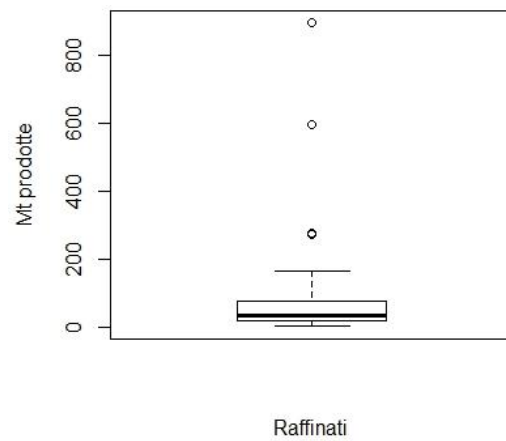
#### Coefficiente di variazione

p.greggio	2.05
p.raffinati	2.05
p.coalignite	3.36
p.gas	2.39
p.ele	2.41
co2	2.5

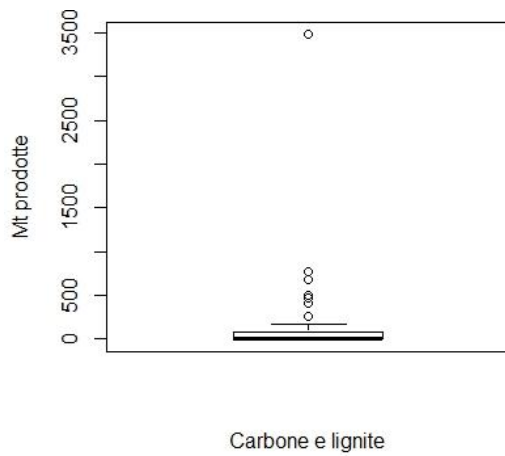
boxplot 1: Produzione di greggio



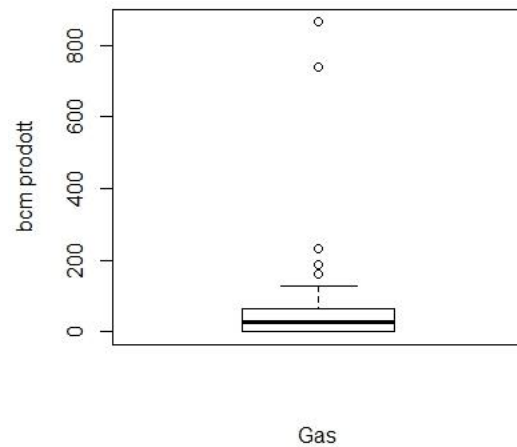
boxplot 2: Produzione di raffinati



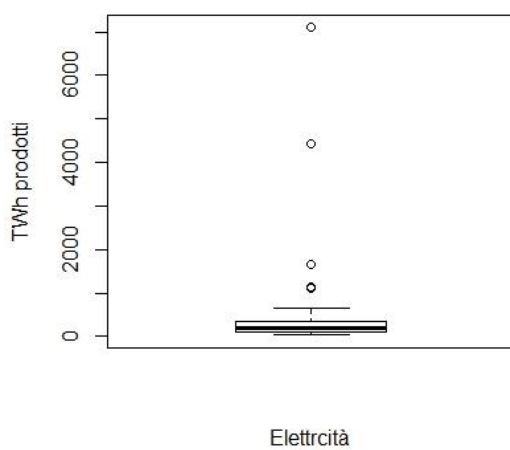
boxplot 3: Produzione di carbone e lignite



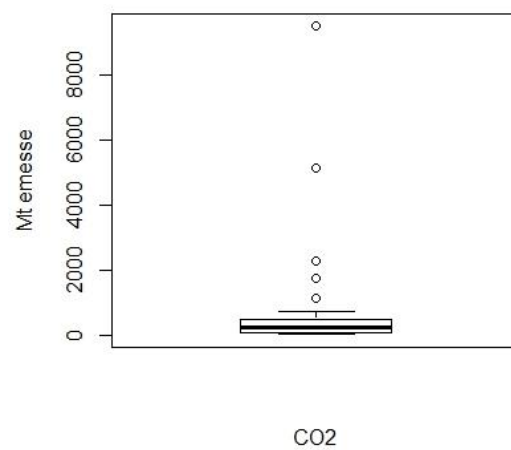
boxplot 4: Produzione di gas



boxplot 5: Produzione di energia elettrica

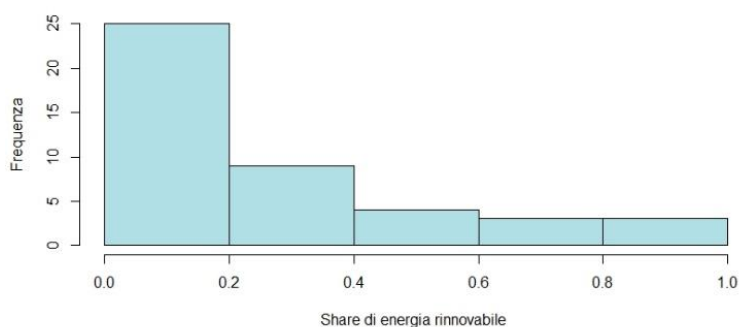


boxplot 6: Produzione di CO2



## Share di energia rinnovabile, quali sono i paesi con uno share superiore al 65%?

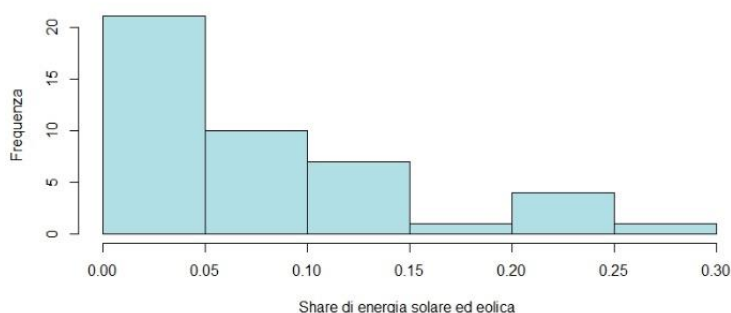
Istogramma 1: Share di energia rinnovabile



Paese	% energia rinnovabile
Norway	98%
Canada	66%
Brazil	82%
Colombia	76%
Venezuela	71%
New Zealand	83%

## Share di energia solare ed eolica, quali sono i paesi con uno share superiore al 20%?

Istogramma 2: Share di energia solare ed eolica



Paese	% energia solare, eolica
Germany	25%
Portugal	23%
Spain	23%
United Kingdom	21%
New Zealand	22%

## Analisi congiunte (a coppie):

### Connessione tra variabili categoriche

### Ripartizione dei Paesi per livelli di CO<sub>2</sub>

	basso	medio	elevato	
africa	4,5%	4,5%	0,0%	9,1%
americas	9,1%	6,8%	2,3%	18,2%
asia	9,1%	22,7%	4,5%	36,4%
europa	18,2%	15,9%	2,3%	36,4%
	40,9%	50,0%	9,1%	100,0%

	basso	medio	elevato	
africa	2	2	0	4
americas	4	3	1	8
asia	4	10	2	16
europa	8	7	1	16
	18	22	4	44

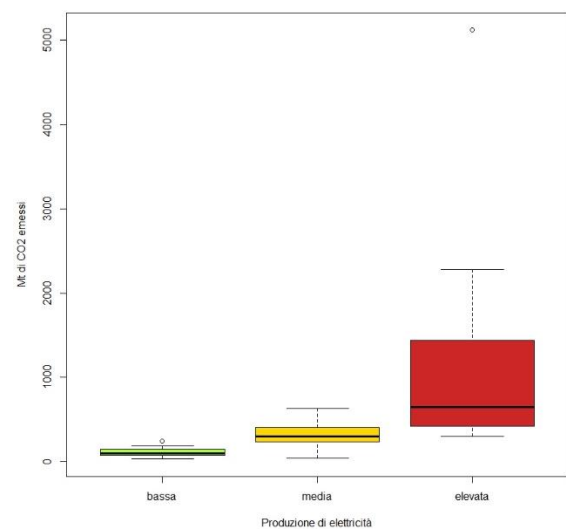
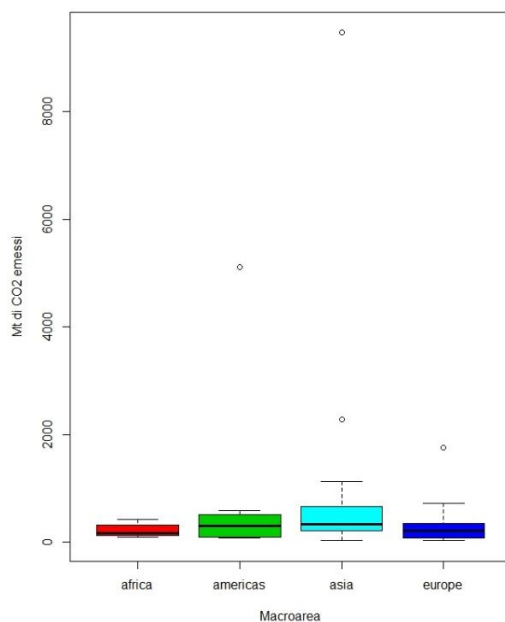
Le parziali e le marginali di colonna permettono di comprendere in che percentuale ciascuna macroarea contribuisce alle emissioni di CO<sub>2</sub>. Inoltre, la V di Cramer [0,1] che misura il grado di connessione tra le due variabili categoriche è 0,19. Si può affermare quindi che tra le due variabili non ci sia una debole connessione.

### Connessione tra produzione di energia elettrica (righe) e livello di emissione di CO<sub>2</sub> (colonne)

	basso	medio	elevato	
bassa	32,6%	2,3%	0,0%	34,9%
media	9,3%	25,6%	0,0%	34,9%
elevata	0,0%	23,3%	7,0%	30,2%
	41,9%	51,2%	7,0%	100,0%

	basso	medio	elevato	
bassa	14	1	0	15
media	4	11	0	15
elevata	0	10	3	13
	18	22	3	43

In questo caso il valore della V di Cramer è 0,59. La connessione tra livelli di produzione elettrica e livelli di emissione di CO<sub>2</sub> è considerevole.



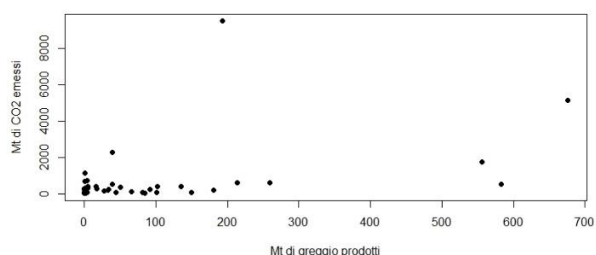
I boxplot sopra rappresentati considerano le stesse variabili delle tabelle senza suddividere in categorie l'emissione di CO<sub>2</sub> rendendo visibilmente chiara la dipendenza tra le variabili.

## Correlazione tra variabili

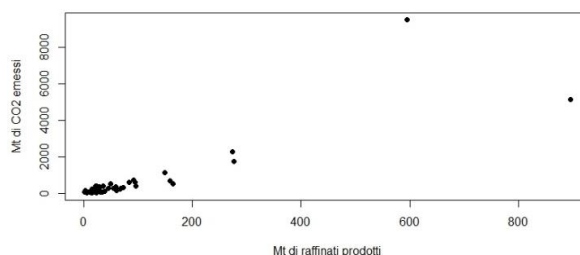
	p.greggio	p.raffinati	p.coalignite	p.gas	p.ele	co2
p.greggio	1	0,678	0,222	0,833	0,44	0,431
p.raffinati	0,678	1	0,645	0,742	0,886	0,861
p.coalignite	0,222	0,645	1	0,261	0,902	0,933
p.gas	0,833	0,742	0,261	1	0,493	0,479
p.ele	0,44	0,886	0,902	0,493	1	0,994
co2	0,431	0,861	0,933	0,479	0,994	1

È possibile notare come la produzione di energia elettrica sia quasi perfettamente correlata all'emissione di CO<sub>2</sub>. L'energia elettrica generata da fonti rinnovabili rappresenta infatti solo il 28% dell'energia elettrica totale, il restante è costituito dalla combustione di combustibili fossili come carbone, lignite, petrolio, idrocarburi vari, gas di scisto e altri gas naturali.

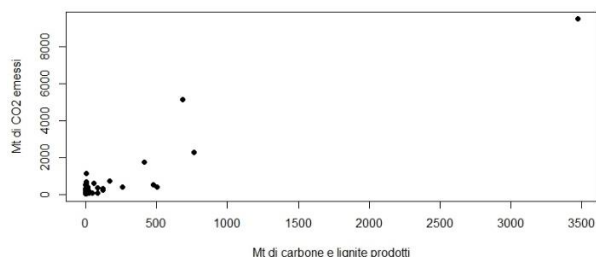
Plot 1:  $r=0,431$



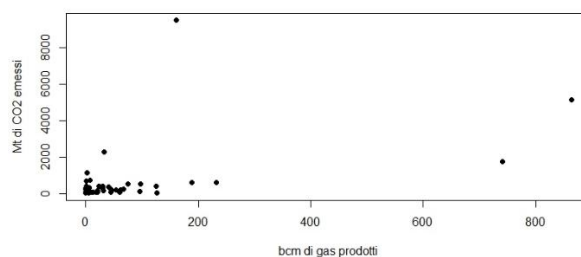
Plot 2:  $r=0,861$



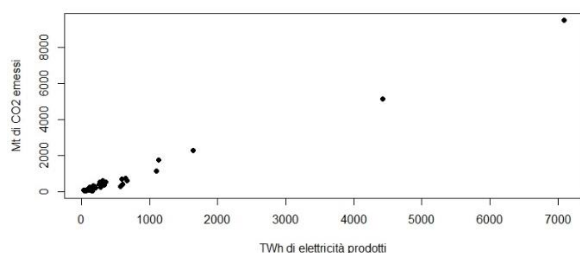
Plot 3:  $r=0,933$



Plot 4:  $r=0,479$

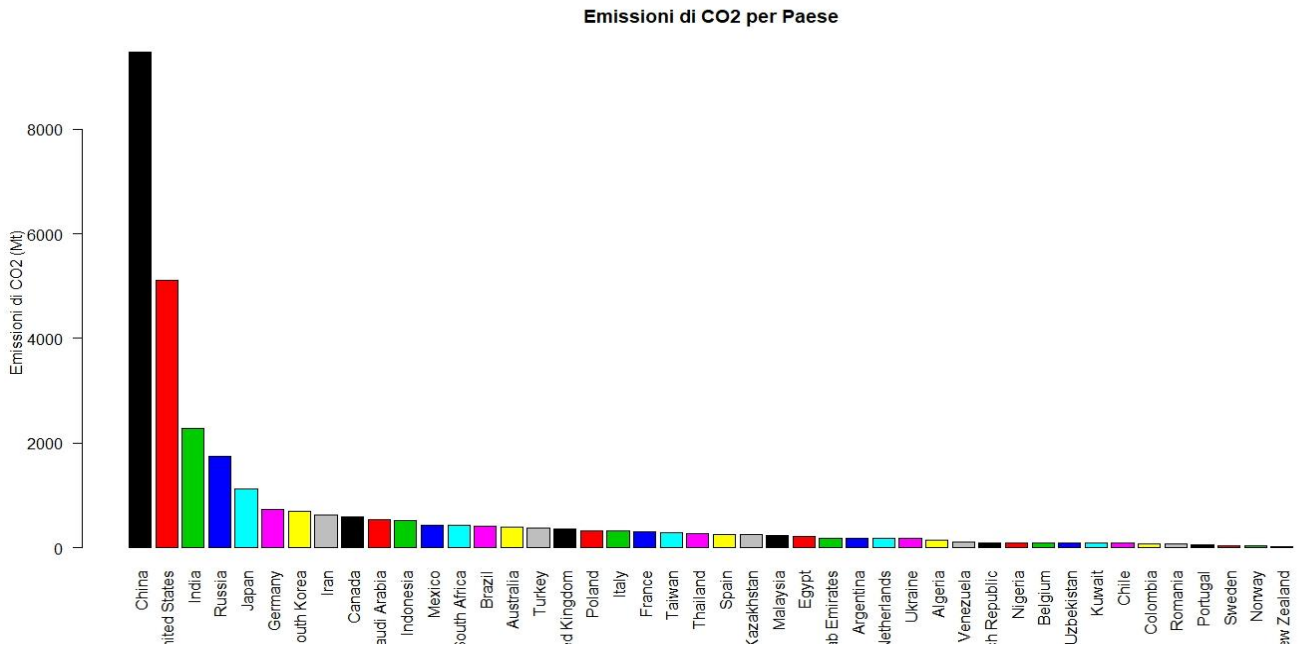


Plot 5:  $r=0,994$



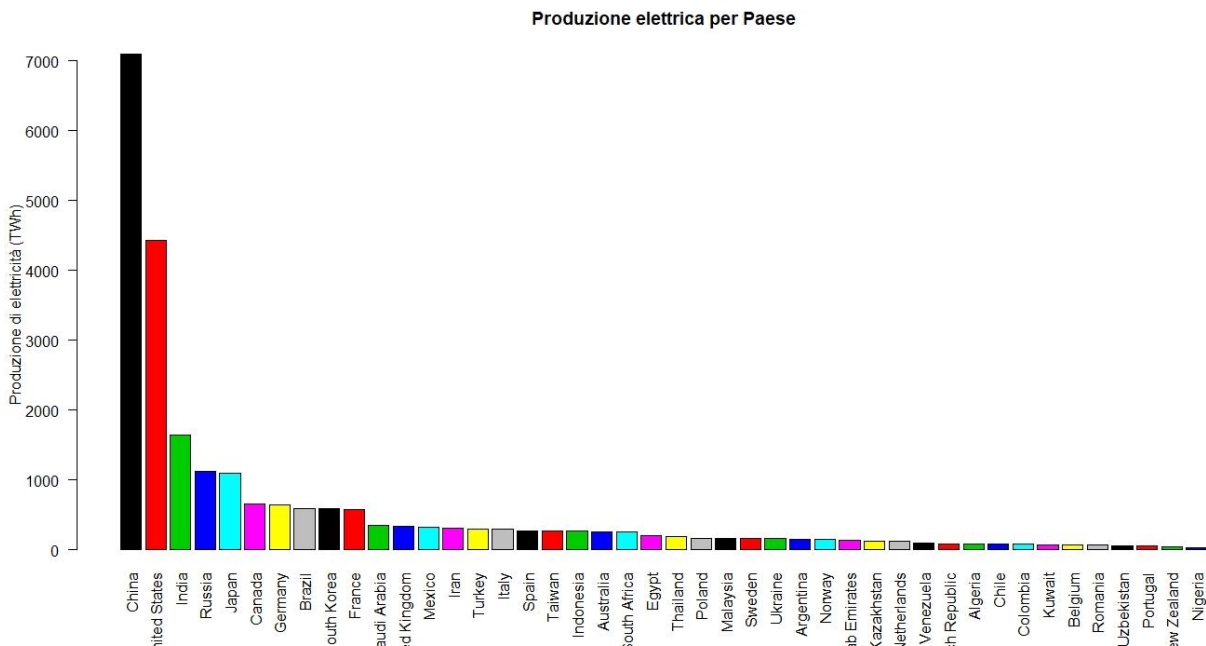


## Emissioni di CO<sub>2</sub>



Il grafico a barre mostra come gran parte della CO<sub>2</sub> nel 2018 è emessa dalla Cina e dagli Stati Uniti che insieme all'economia in via di sviluppo dell'India producono 16861,22 Mt su 30196,40 Mt totali, ovvero il 55,83%.

## Produzione di elettricità



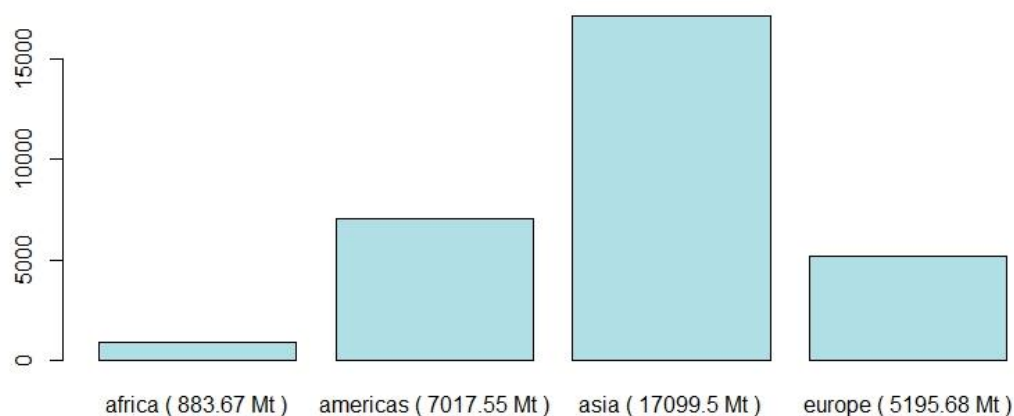
Paesi	United States	China	India
co2.top3	5117,77	9466,5	2276,95
ele.top3	4429,26	7091,79	1643,17

Si può notare come i tre Paesi più impattanti siano anche quelli che producono più energia e quelli con una popolazione decisamente superiore:

la sola Cina nel 2018 contava 1,43 miliardi di abitanti (18,79%), l'India 1,35 miliardi (17,72%) e gli Stati Uniti d'America 327,09 milioni (4,29%) a fronte dei 7,63 miliardi totali.

Fonte dati: <https://www.populationpyramid.net/it/>

**Emissioni di CO2 per ciascuna macroarea geografica**



Macroaree	africa	americas	asia	europe	
Mt di CO2 emessa	883,67	7017,55	17099,5	5195,68	30196,4
% rispetto al totale	3%	23%	57%	17%	100%

Il grafico a barre e la tabella ci mostrano in che misura ciascuna macroarea contribuisce complessivamente alle emissioni di CO<sub>2</sub>, è possibile notare come la sola Asia rappresenti il 57% delle emissioni totali (2018) nel mondo.

In conclusione si può affermare che, i livelli di CO<sub>2</sub> sono notevolmente alti nei Paesi in cui una maggiore popolazione richiede un'elevata quantità di energia per il sostentamento; vi sono Paesi in cui la percentuale di energia ricavata da fonti rinnovabili è molto alta e ciò oltre ad essere un fattore positivo, può anche essere d'esempio a realtà in via di sviluppo come la Cina e l'India. Proprio in Asia infatti è dove il 57% percento delle 30196 Mt di CO<sub>2</sub> nel 2018 sono state emesse.

---

Giulio Mazzarelli

Fine