

Progetto Social Media Analytics: “Artificial Intelligence”

-FEDERICO DE LUCA

-JACOPO MARTINES

Introduzione:

L'intelligenza artificiale (AI) rappresenta una delle tecnologie più rivoluzionarie e trasformative del nostro tempo, capace di ridefinire modelli di business, processi produttivi e strategie aziendali. Questo lavoro si propone di analizzare come le imprese europee, suddivise per classi dimensionali, stiano adottando l'AI e quali fattori ne influenzino l'adozione.

Il primo capitolo fornisce un quadro teorico iniziale, introducendo lo stato dell'arte sull'intelligenza artificiale e analizzando i principali aspetti tecnologici, economici e sociali legati alla sua implementazione nelle aziende. Si affrontano sia le opportunità che le sfide connesse all'utilizzo dell'AI, delineando un contesto utile per le successive analisi.

Il secondo capitolo descrive l'adozione dell'AI nelle imprese europee attraverso un'analisi basata su dati Eurostat. In particolare, si esamina la diffusione delle tecnologie di AI in base alle classi dimensionali delle aziende, evidenziando barriere come la mancanza di chiarezza sulle implicazioni legali. L'analisi descrittiva è supportata da grafici che illustrano tendenze e gap tecnologici tra i paesi dell'Unione Europea.

Il terzo capitolo si concentra sull'analisi testuale di articoli accademici e contributi scientifici presenti in letteratura, esplorando come l'AI venga discussa in termini di applicazioni, impatti e prospettive future. Attraverso questa analisi, si cerca di estrapolare temi ricorrenti e modelli interpretativi utili per comprendere le dinamiche in atto.

Infine, il quarto capitolo presenta un'analisi di commenti e opinioni raccolti tramite scraping su Twitter (X), effettuato con l'utilizzo di API specifiche. Questa parte del lavoro si propone di esplorare le percezioni degli utenti, identificando sentimenti prevalenti, trend emergenti e differenze di opinione sul tema dell'AI, con un focus sui commenti rilevanti per il contesto europeo.

L'obiettivo complessivo è fornire una visione integrata dell'adozione dell'intelligenza artificiale nelle aziende europee, combinando analisi descrittive e testuali con dati provenienti sia da fonti ufficiali sia dai social media. Questo approccio multidimensionale mira a offrire spunti di riflessione per favorire un'adozione più inclusiva e consapevole dell'AI, contribuendo a ridurre le disuguaglianze tra operatori economici di diverse dimensioni.

CAPITOLO 1: Origini e sviluppo dell'intelligenza artificiale

Il termine “intelligenza artificiale” (IA) è stato coniato nel 1956 durante il celebre workshop di Dartmouth, considerato il punto di partenza ufficiale di questa disciplina. Il concetto, tuttavia, trae le sue origini da precedenti riflessioni filosofiche e scientifiche sull'automazione e sulla capacità delle macchine di imitare l'intelligenza umana. Tra i precursori si possono citare Alan Turing, che nel 1950 introdusse il celebre “Test di Turing”, e John McCarthy, che contribuì alla formalizzazione dell'IA come campo accademico.

Negli anni '60 e '70, l'IA ha registrato significativi progressi grazie all'elaborazione di algoritmi di ricerca e al primo sviluppo di linguaggi di programmazione come Lisp. Tuttavia, il periodo conosciuto come “AI Winter” negli anni '80 ha rallentato gli investimenti nel settore, a causa di aspettative non soddisfatte.

Con l'avvento di hardware più potenti e l'accesso a grandi quantità di dati negli anni 2000, l'IA ha vissuto una nuova era di crescita. Tecnologie come il machine learning, il deep learning e le reti neurali convoluzionali hanno rivoluzionato il settore, con applicazioni che spaziano dalla medicina alla finanza, dai trasporti alla sicurezza informatica.

Attualmente, l'intelligenza artificiale rappresenta una delle tecnologie più promettenti del nostro tempo, con progressi costanti verso sistemi sempre più autonomi e intelligenti. Nonostante le straordinarie opportunità offerte, permangono numerose sfide, tra cui questioni etiche, bias algoritmici e l'impatto sull'occupazione.

BIBLIOGRAFIA

- Simon, H. (2009) Neural networks and learning machines.
- Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & Data, M. (2005) Practical machine learning tools and techniques.
- Jiawei, H., & Micheline, K. (2006) Data mining: concepts and techniques.
- Sharma, A., and Pandey, H. (2020) Big Data and Analytics in Industry 4.0, Springer Nature Switzerland AG

CAPITOLO 2: Analisi descrittiva

L'analisi descrittiva condotta si basa sui dati Eurostat disponibili tramite il dataset “Artificial Intelligence by size class of enterprise”, il quale offre una panoramica sull'utilizzo dell'intelligenza artificiale da parte delle imprese europee, suddivise per classe dimensionale e per stato. Le aziende sono classificate in tre categorie:

- Piccole (10-49 dipendenti)
- Medie (50-249 dipendenti)
- Grandi (oltre 250 dipendenti).

Sono state escluse dall'analisi le aziende dei settori agricolo, ittico e minerario, nonché il settore finanziario. L'unità di misura utilizzata è la percentuale di imprese. Le informazioni sono relative a diverse domande specifiche, per un totale di 77 indicatori.

Online data code: isoc_eb_ai

2.1 Analisi del primo sotto-dataset: Imprese che utilizzano AI (classe 10-49 dipendenti)

Information society indicator: Enterprises don't use any AI system (of E_CHTB, E_BDAML, E_BDANL, E_RBTS).

- [E_BDAML]: Analyse big data internally using machine learning
- [E_BDANL]: Analyse big data internally using natural language processing, natural language generation or speech recognition
- [E_RBTS] Use service robots
- [E_CHTB] Enterprises with a chat service where a chatbot or a virtual agent replies to customers
- [E_AI_0] Enterprises don't use any AI system (of E_CHTB, E_BDAML, E_BDANL, E_RBTS)

Unit of measure: Percentage of enterprises

Nel primo sotto-dataset analizzato, l'indicatore che avevamo riguardava le piccole imprese (10-49 dipendenti) che non utilizzano alcun sistema di intelligenza artificiale nel 2020. Quello che abbiamo utilizzato però è stato il suo complemento a 100 ossia le imprese di questa dimensione che utilizzano almeno una delle precedenti tecnologie AI.

Il dataset è stato preprocessato e visualizzato tramite un barplot che rappresenta la distribuzione delle percentuali per ogni nazione europea.

Il tutto è stato ripetuto per le aziende aventi classi di dimensioni diverse.

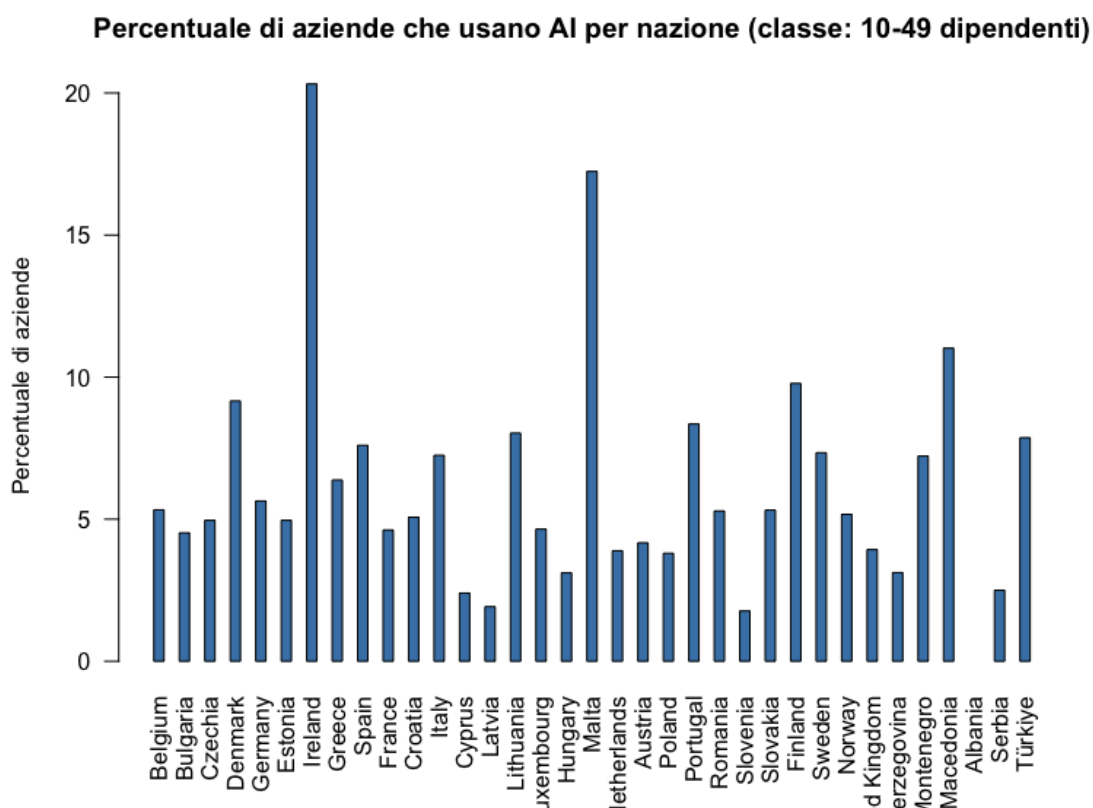
```
library(readxl)
library(DescTools)
dataset_10_49_E_AI_0 <-
read_excel("isoc_eb_ai__custom_14777017_page_spreadsheet.xlsx")

## New names:
## • `` -> `...3`
## • `` -> `...5`
## • `` -> `...7`
## • `` -> `...9`
## • `` -> `...11`

dataset_10_49_E_AI_0 <- dataset_10_49_E_AI_0[-1,]
dataset_10_49_E_AI_0 <- dataset_10_49_E_AI_0[,1:2]
dataset_10_49_E_AI_0$'2020' <- as.numeric(dataset_10_49_E_AI_0$'2020')
```

```
## Warning: NAs introduced by coercion
```

```
dataset_10_49_E_AI_0$"2020_AI_SI" <- 100 - dataset_10_49_E_AI_0$"2020"  
barplot(  
  height = dataset_10_49_E_AI_0$"2020_AI_SI",  
  space = 1.5,  
  names.arg = dataset_10_49_E_AI_0$"TIME",  
  col = "steelblue",  
  main = "Percentuale di aziende che usano AI per nazione (classe: 10-49  
dipendenti)",  
  ylab = "Percentuale di aziende",  
  las = 2 # Ruota le etichette sull'asse X  
)
```



2.1.1 Analisi del secondo sotto-dataset: Imprese che utilizzano AI (classe 50-249 dipendenti)

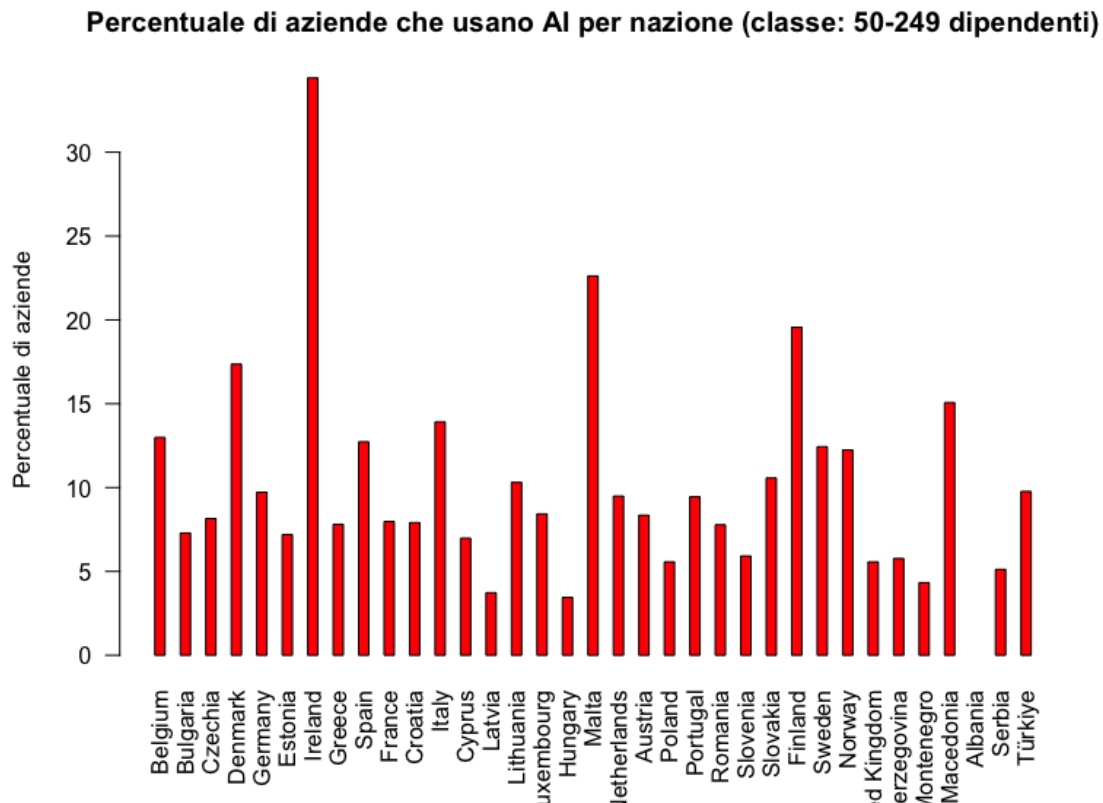
```
library(readxl)  
dataset_50_249_E_AI_0 <-  
read_excel("isoc_eb_ai__custom_14777150_page_spreadsheet.xlsx")  
  
## New names:  
## • `` -> `...3`  
## • `` -> `...5`  
## • `` -> `...7`
```

```
## • `` -> `...9`
## • `` -> `...11`

dataset_50_249_E_AI_0 <- dataset_50_249_E_AI_0[-1,]
dataset_50_249_E_AI_0 <- dataset_50_249_E_AI_0[,1:2]
dataset_50_249_E_AI_0$'2020' <- as.numeric(dataset_50_249_E_AI_0$'2020')

## Warning: NAs introduced by coercion

dataset_50_249_E_AI_0$"2020_AI_SI" <- 100 - dataset_50_249_E_AI_0$'2020'
barplot(
  height = dataset_50_249_E_AI_0$'2020_AI_SI',
  space = 1.5,
  names.arg = dataset_50_249_E_AI_0$'TIME',
  col = "red",
  main = "Percentuale di aziende che usano AI per nazione (classe: 50-249
dipendenti)",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota Le etichette sull'asse X
)
```



2.1.2 Analisi del terzo sotto-dataset: Imprese che utilizzano AI (oltre 250 dipendenti)

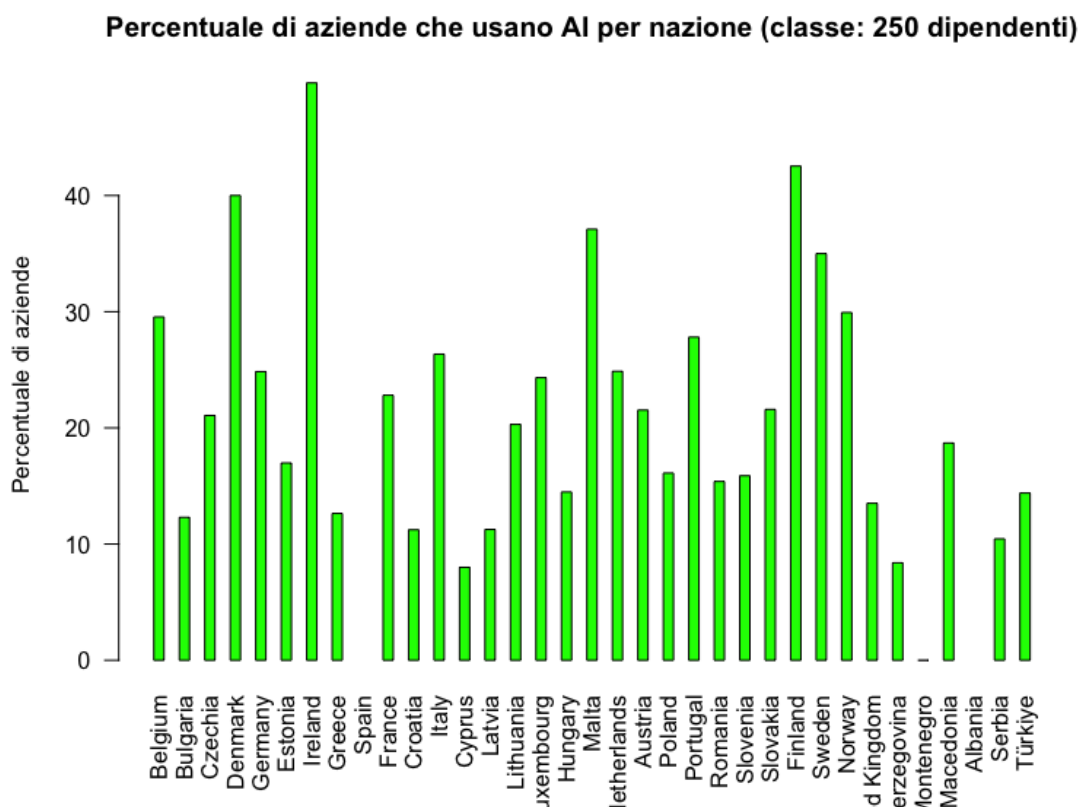
```
dataset_250_E_AI_0 <- read_excel("E_AI_0_250.xlsx")

## New names:
## • `` -> `...3`

dataset_250_E_AI_0 <- dataset_250_E_AI_0[-1,]
dataset_250_E_AI_0 <- dataset_250_E_AI_0[,1:2]
dataset_250_E_AI_0$'2020' <- as.numeric(dataset_250_E_AI_0$'2020')

## Warning: NAs introduced by coercion

dataset_250_E_AI_0$'2020_AI_SI' <- 100 - dataset_250_E_AI_0$'2020'
barplot(
  height = dataset_250_E_AI_0 $'2020_AI_SI',
  space = 1.5,
  names.arg = dataset_250_E_AI_0 $'TIME',
  col = "green",
  main = "Percentuale di aziende che usano AI per nazione (classe: 250
dipendenti)",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota Le etichette sull'asse X
)
```



2.2 Analisi Specifica: Imprese che Utilizzano Tecnologie di AI per Ricerca, Sviluppo e Innovazione

Online Code: E_AI_PRDI

Un ulteriore approfondimento è stato condotto sulle imprese che utilizzano tecnologie di intelligenza artificiale per attività di ricerca e sviluppo (R&D) o innovazione (domanda E_AI_PRDI). L'analisi include i dati disponibili per gli anni 2023 e 2024, considerando tutte le aziende di vari paesi. Sono stati generati barplot annuali per evidenziare le variazioni tra i diversi Stati. Inoltre, è stata fornita una descrizione statistica riassuntiva per esplorare le tendenze e le differenze nell'adozione dell'AI in ambito R&D e innovazione tra le varie nazioni.

```
dataset_E_AI_PRDI <- read_excel("isoc_eb_ai_E_AI_PRDI.xlsx")

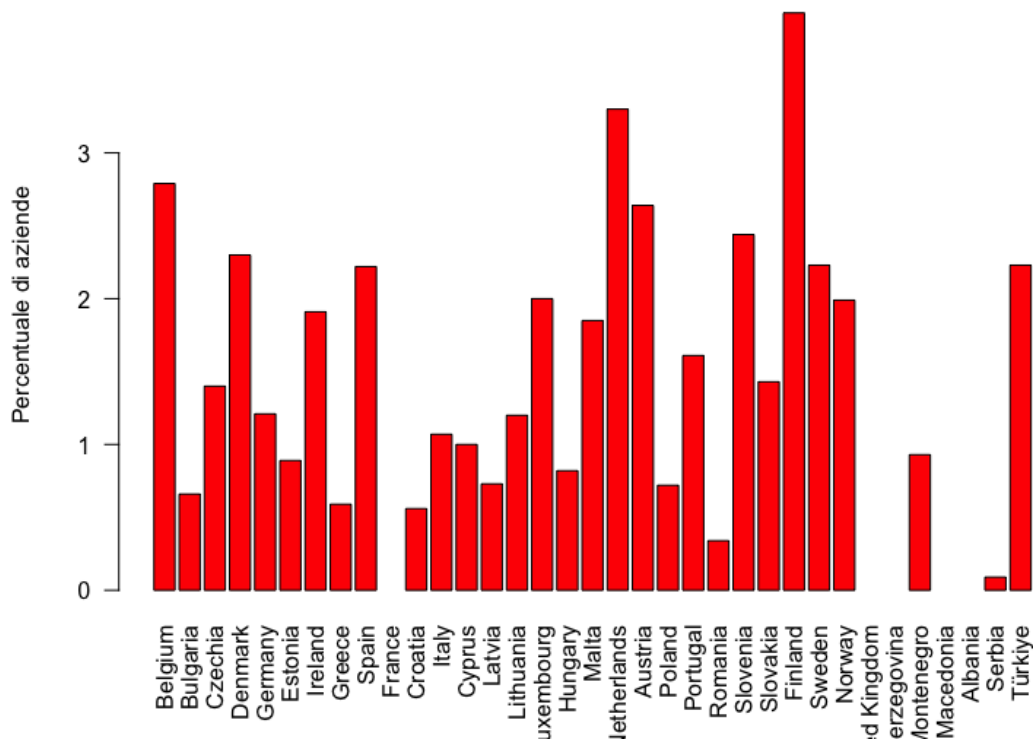
dataset_E_AI_PRDI[, c(2, 3)] <- lapply(dataset_E_AI_PRDI[, c(2, 3)], as.numeric)

## Warning in lapply(dataset_E_AI_PRDI[, c(2, 3)], as.numeric): NAs introduced by
## coercion
## Warning in lapply(dataset_E_AI_PRDI[, c(2, 3)], as.numeric): NAs introduced by
## coercion

dataset_E_AI_PRDI <- dataset_E_AI_PRDI[-1,]
colnames(dataset_E_AI_PRDI)[1] <- "GEO_label"

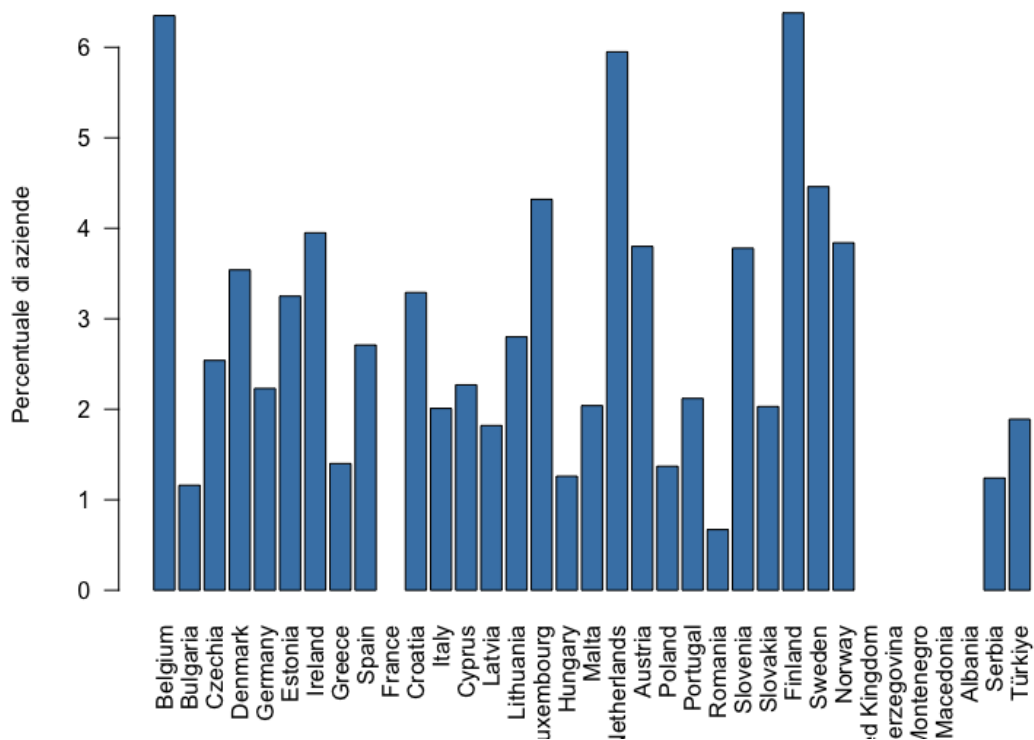
barplot(
  height = dataset_E_AI_PRDI$'2023',
  names.arg = dataset_E_AI_PRDI$GEO_label,
  col = "red",
  main = "Percentuale di aziende che usano l'AI per la ricerca e sviluppo 2023",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)
```

Percentuale di aziende che usano l'AI per la ricerca e sviluppo 2023



```
barplot(
  height = dataset_E_AI_PRDI$'2024',
  names.arg = dataset_E_AI_PRDI$GEO_label,
  col = "steelblue",
  main = "Percentuale di aziende che usano l'AI per la ricerca e sviluppo 2024",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)
```


Percentuale di aziende che usano l'AI per la ricerca e sviluppo 2024

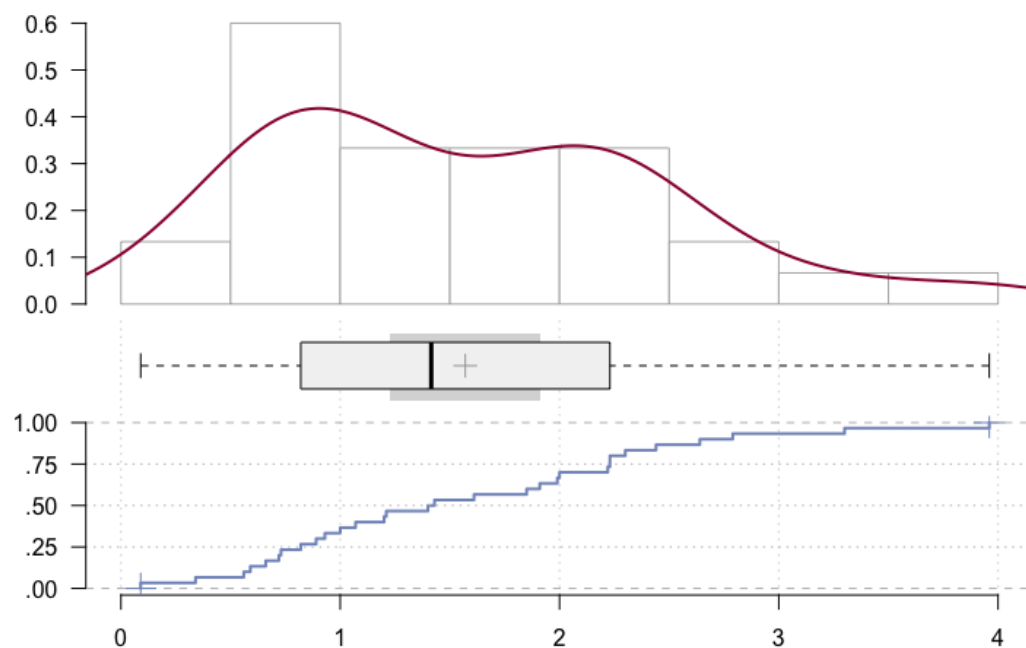


```
Desc(dataset_E_AI_PRDI[,2:3])

##
## Describe dataset_E_AI_PRDI[, 2:3] (tbl_df, tbl, data.frame):
##
## data frame: 35 obs. of 2 variables
## 29 complete cases (82.9%)
##
##   Nr  Class ColName  NAs      Levels
##   1   num   2023     5 (14.3%)
##   2   num   2024     6 (17.1%)
##
##
## 1 - 2023 (numeric)
##
##   length      n    NAs  unique      0s    mean  meanCI'
##      35      30      5      29      0  1.5703  1.2254
##           85.7%  14.3%           0.0%           1.9152
##
##   .05   .10   .25 median   .75   .90   .95
## 0.4390 0.5870 0.8375 1.4150 2.2275 2.6550 3.0705
##
##   range      sd  vcoef      mad      IQR      skew      kurt
## 3.8700 0.9237 0.5882 1.0230 1.3900 0.5714 -0.3327
##
```

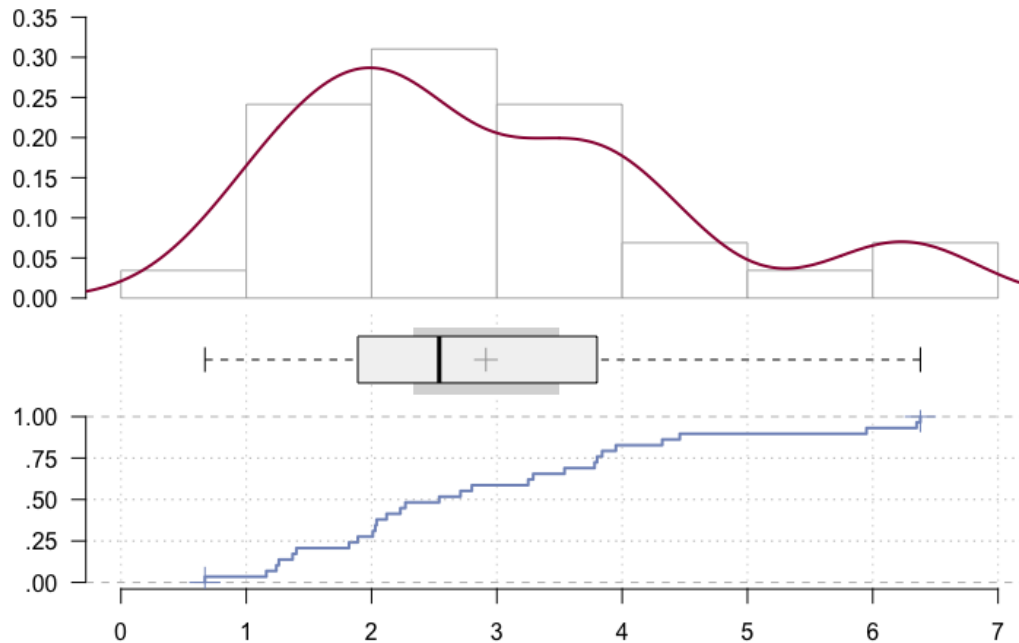
```
## lowest : 0.09, 0.34, 0.56, 0.59, 0.66
## highest: 2.44, 2.64, 2.79, 3.3, 3.96
##
## ' 95%-CI (classic)
```

1 - 2023 (numeric)



```
##
## 2 - 2024 (numeric)
##
## length      n      NAs  unique      0s      mean  meanCI '
##      35      29      6      = n      0      2.913  2.328
##      82.9%  17.1%      0.0%      3.497
##
##      .05     .10     .25  median     .75     .90     .95
##      1.192  1.256  1.890  2.540    3.800  4.758  6.190
##
##      range     sd  vcoef     mad     IQR     skew     kurt
##      5.710    1.537  0.528    1.690    1.910    0.777   -0.231
##
## lowest : 0.67, 1.16, 1.24, 1.26, 1.37
## highest: 4.32, 4.46, 5.95, 6.35, 6.38
##
## ' 95%-CI (classic)
```

2 - 2024 (numeric)



2.3 Analisi Specifica: Imprese che Utilizzano Tecnologie di AI per i Processi Produttivi

Online Code: #E_AI_PPP

Un ulteriore approfondimento è stato condotto sulle imprese che utilizzano tecnologie di intelligenza artificiale per i processi produttivi (domanda E_AI_PPP). L'analisi include i dati disponibili per gli anni 2021, 2023 e 2024, considerando aziende di diverse dimensioni e paesi. Sono stati generati barplot annuali per evidenziare le variazioni nell'adozione dell'AI nei processi produttivi tra i vari Stati. Inoltre, è stata fornita una descrizione statistica riassuntiva per analizzare l'evoluzione dell'uso delle tecnologie di AI nei settori produttivi nel corso degli anni.

```
dataset_E_AI_PPP <- read_excel("isoc_eb_ai_E_AI_PPP.xlsx")

dataset_E_AI_PPP[, c(2, 3,4)] <- lapply(dataset_E_AI_PPP[, c(2, 3,4)], as.numeric)

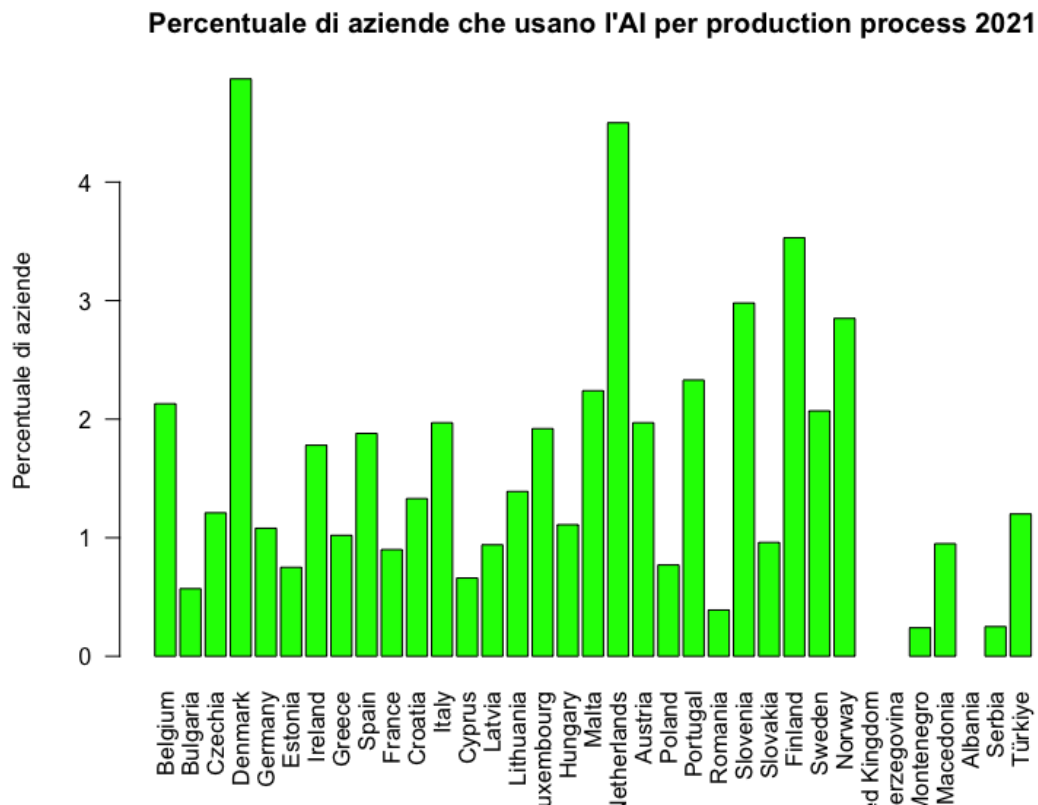
## Warning in lapply(dataset_E_AI_PPP[, c(2, 3, 4)], as.numeric): NAs introduced
## by coercion
## Warning in lapply(dataset_E_AI_PPP[, c(2, 3, 4)], as.numeric): NAs introduced
## by coercion
## Warning in lapply(dataset_E_AI_PPP[, c(2, 3, 4)], as.numeric): NAs introduced
## by coercion

dataset_E_AI_PPP <- dataset_E_AI_PPP[-1,]
colnames(dataset_E_AI_PPP)[1] <- "GEO_label"
```

```

barplot(
  height = dataset_E_AI_PPP$'2021',
  names.arg = dataset_E_AI_PPP$GEO_label,
  col = "green",
  main = "Percentuale di aziende che usano l'AI per production process 2021",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)

```

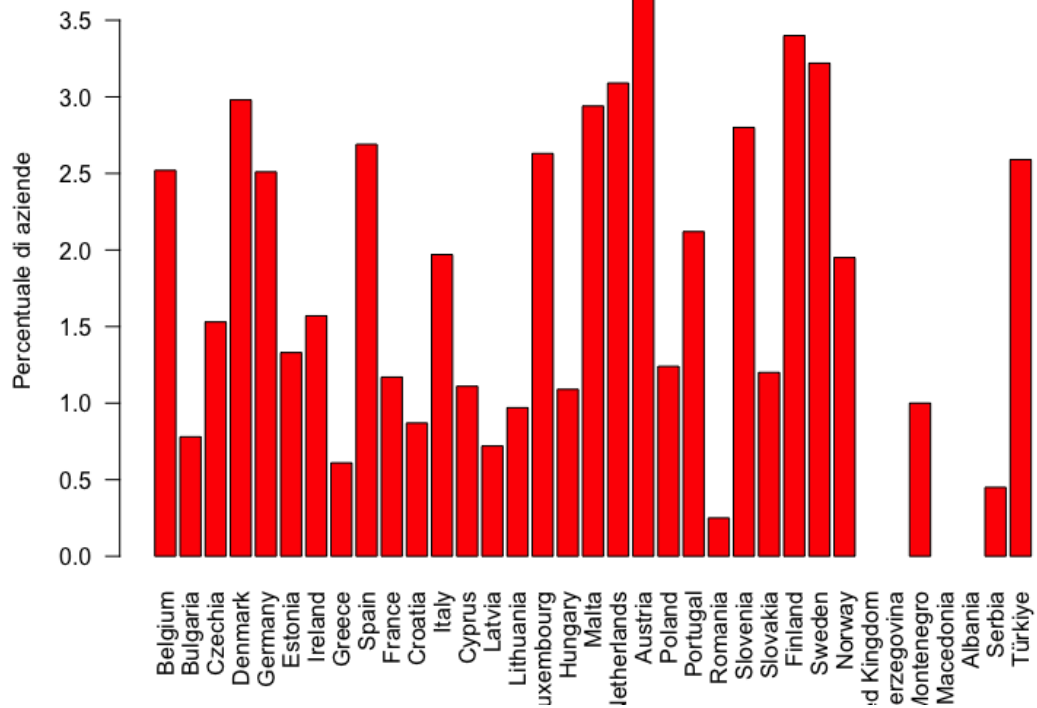


```

barplot(
  height = dataset_E_AI_PPP$'2023',
  names.arg = dataset_E_AI_PPP$GEO_label,
  col = "red",
  main = "Percentuale di aziende che usano l'AI per production process 2023",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)

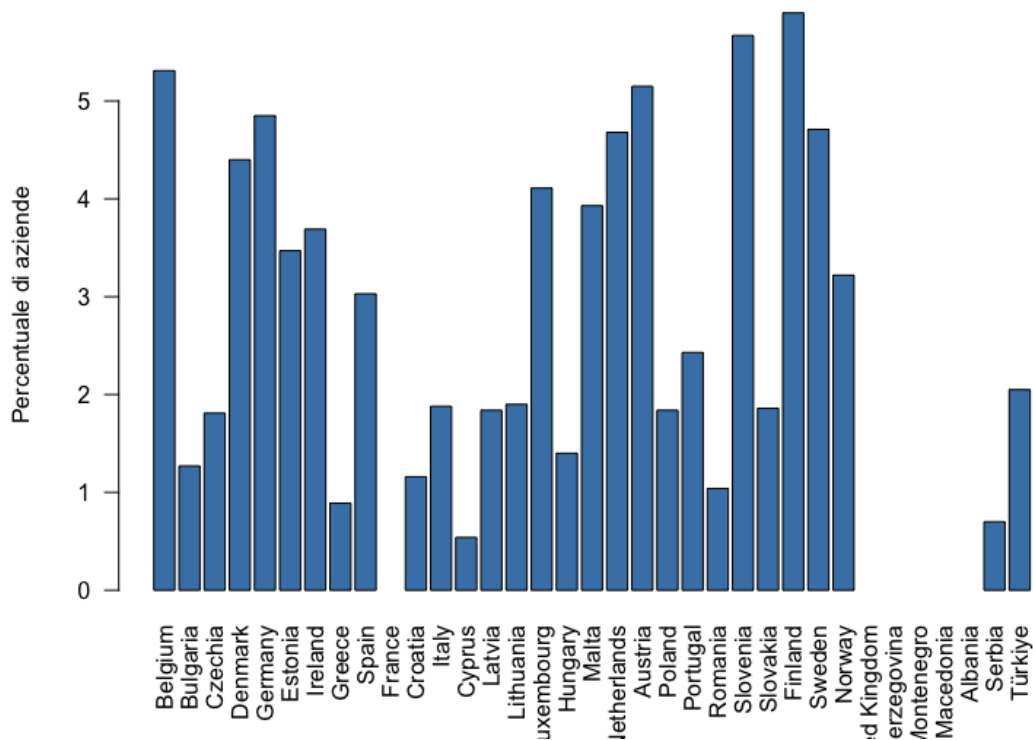
```

Percentuale di aziende che usano l'AI per production process 2023



```
barplot(
  height = dataset_E_AI_PPP$'2024',
  names.arg = dataset_E_AI_PPP$GEO_label,
  col = "steelblue",
  main = "Percentuale di aziende che usano l'AI per production process 2024",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)
```

Percentuale di aziende che usano l'AI per production process 2024

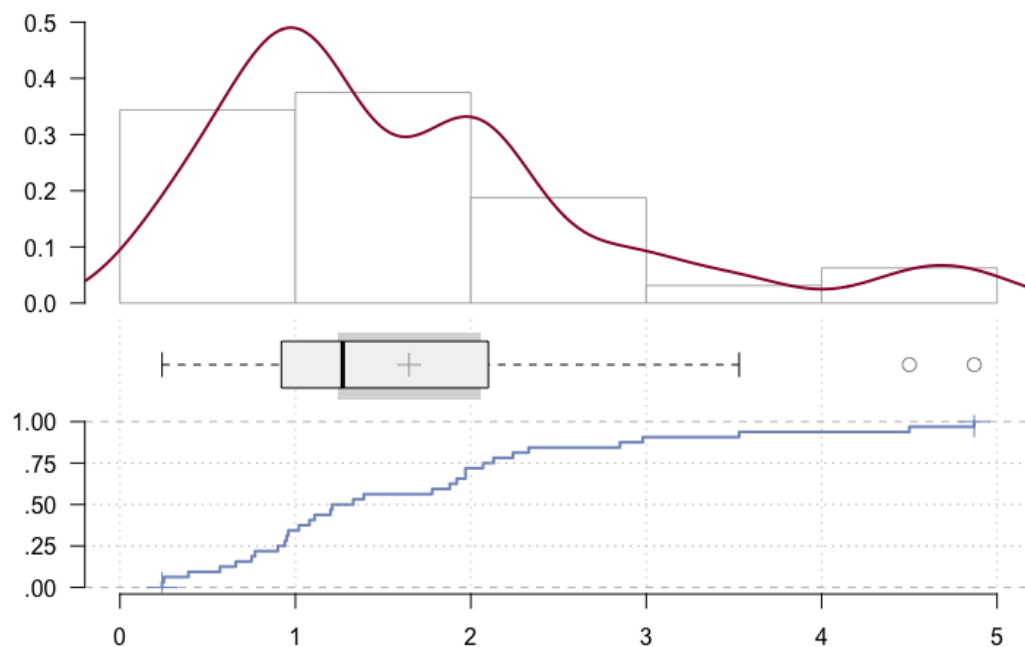


```
Desc(dataset_E_AI_PPP[,2:4])

## -----
## Describe dataset_E_AI_PPP[, 2:4] (tbl_df, tbl, data.frame):
##
## data frame: 35 obs. of 3 variables
##      29 complete cases (82.9%)
##
##   Nr  Class ColName  NAs      Levels
##   1   num   2021     3 (8.6%)
##   2   num   2023     4 (11.4%)
##   3   num   2024     6 (17.1%)
##
## -----
## 1 - 2021 (numeric)
##
##   length      n    NAs  unique      0s    mean  meanCI'
##      35      32      3      31      0  1.6481  1.2402
##           91.4%   8.6%           0.0%           2.0560
##
##   .05   .10   .25  median   .75   .90   .95
## 0.3270 0.5790 0.9300 1.2700 2.0850 2.9670 3.9665
##
##   range      sd  vcoef      mad      IQR      skew      kurt
## 4.6300 1.1313 0.6864 0.9340 1.1550 1.1833 0.9564
```

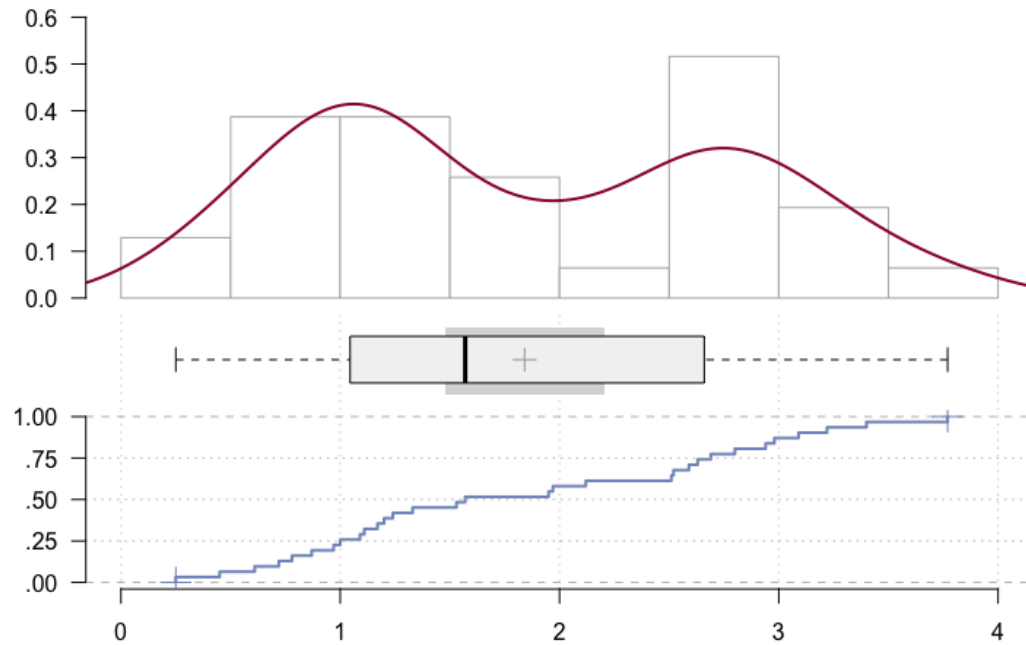
```
##
## lowest : 0.24, 0.25, 0.39, 0.57, 0.66
## highest: 2.85, 2.98, 3.53, 4.5, 4.87
##
## ' 95%-CI (classic)
```

1 - 2021 (numeric)



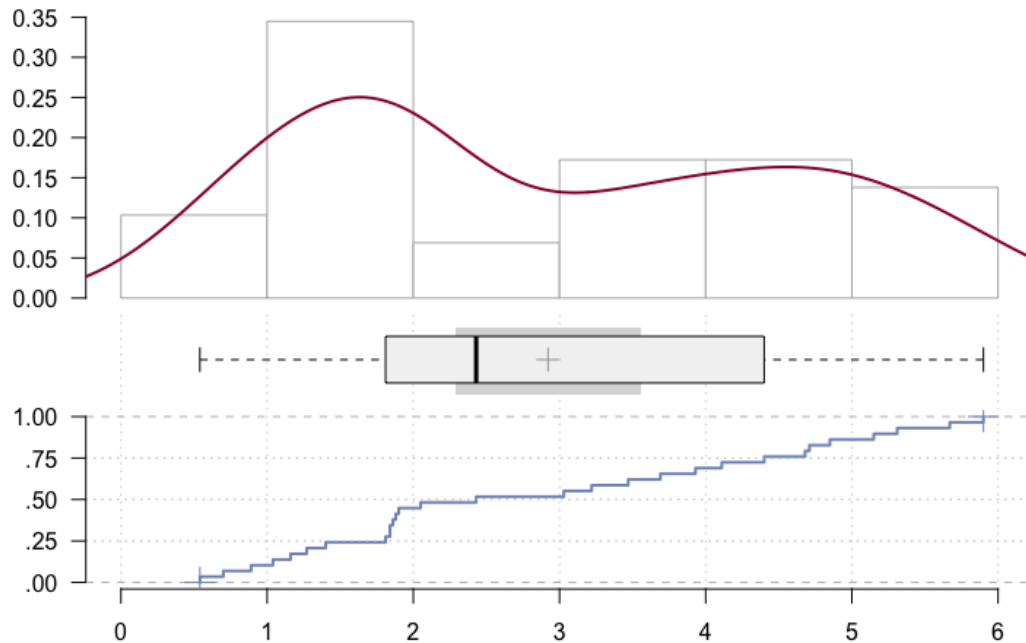
```
##
## 2 - 2023 (numeric)
##
## length      n    NAs  unique      0s    mean  meanCI '
##      35      31     4    = n      0    1.841  1.477
##      88.6%  11.4%      0.0%      2.205
##
##      .05     .10     .25  median    .75    .90    .95
##    0.530  0.720  1.045  1.570  2.660  3.090  3.310
##
## range      sd  vcoef    mad    IQR    skew    kurt
##    3.520  0.991  0.538  1.260  1.615  0.207  -1.334
##
## lowest : 0.25, 0.45, 0.61, 0.72, 0.78
## highest: 2.98, 3.09, 3.22, 3.4, 3.77
##
## ' 95%-CI (classic)
```

2 - 2023 (numeric)



```
##
## 3 - 2024 (numeric)
##
##   length      n    NAs  unique    0s   mean  meanCI '
##      35      29      6     28     0  2.922  2.291
##           82.9%  17.1%           0.0%      3.552
##
##   .05   .10   .25  median   .75   .90   .95
##  0.776  1.010  1.810  2.430  4.400  5.182  5.526
##
##   range     sd  vcoef     mad    IQR   skew   kurt
##   5.360  1.658  0.567   1.883  2.590  0.270 -1.395
##
## lowest : 0.54, 0.7, 0.89, 1.04, 1.16
## highest: 4.85, 5.15, 5.31, 5.67, 5.9
##
## ' 95%-CI (classic)
```


3 - 2024 (numeric)



2.4 Analisi Specifica: Imprese che Utilizzano Tecnologie di AI per Marketing e Vendite

Online code: E_AI_PMS

Un ulteriore approfondimento è stato condotto sulle imprese che utilizzano tecnologie di intelligenza artificiale per attività di marketing e vendite (domanda E_AI_PMS). L'analisi si basa sui dati disponibili per gli anni 2021, 2023 e 2024, considerando aziende di diverse dimensioni e paesi. Sono stati generati barplot annuali per mettere in evidenza le variazioni nell'adozione dell'AI per marketing e vendite nei vari Stati. Inoltre, è stata fornita una descrizione statistica riassuntiva per esaminare l'evoluzione dell'uso delle tecnologie AI nelle attività di marketing e vendite nel corso del tempo.

```
dataset_E_AI_PMS <- read_excel("isoc_eb_ai_E_AI_PMS.xlsx")
```

```
dataset_E_AI_PMS[, c(2, 3, 4)] <- lapply(dataset_E_AI_PMS[, c(2, 3, 4)], as.numeric)
```

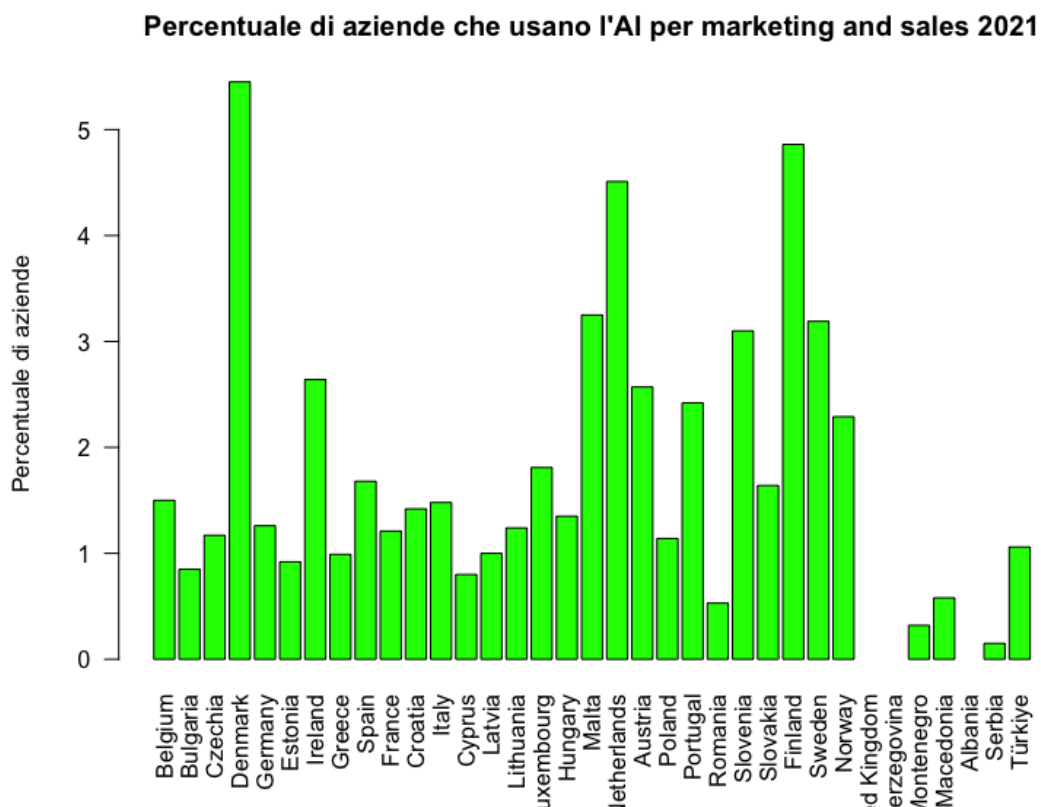
```
## Warning in lapply(dataset_E_AI_PMS[, c(2, 3, 4)], as.numeric): NAs introduced  
## by coercion
```

```
## Warning in lapply(dataset_E_AI_PMS[, c(2, 3, 4)], as.numeric): NAs introduced  
## by coercion
```

```
## Warning in lapply(dataset_E_AI_PMS[, c(2, 3, 4)], as.numeric): NAs introduced
## by coercion
```

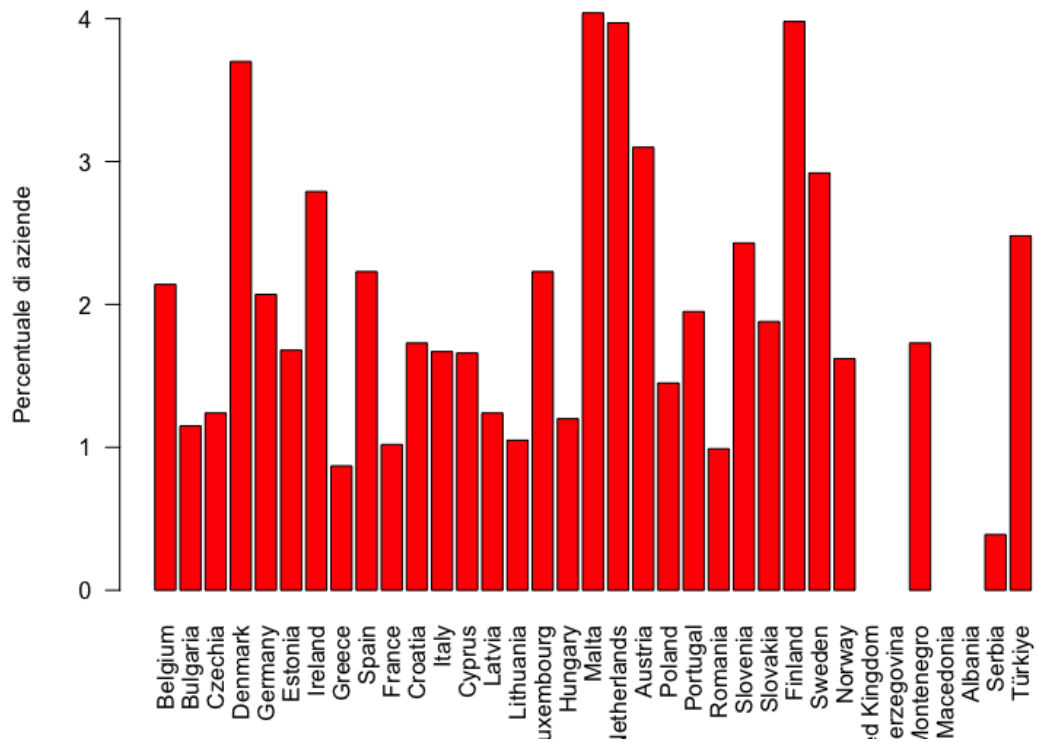
```
dataset_E_AI_PMS <- dataset_E_AI_PMS[-1,]
colnames(dataset_E_AI_PMS)[1] <- "GEO_label"
```

```
barplot(
  height = dataset_E_AI_PMS$'2021',
  names.arg = dataset_E_AI_PMS$GEO_label,
  col = "green",
  main = "Percentuale di aziende che usano l'AI per marketing and sales 2021",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)
```



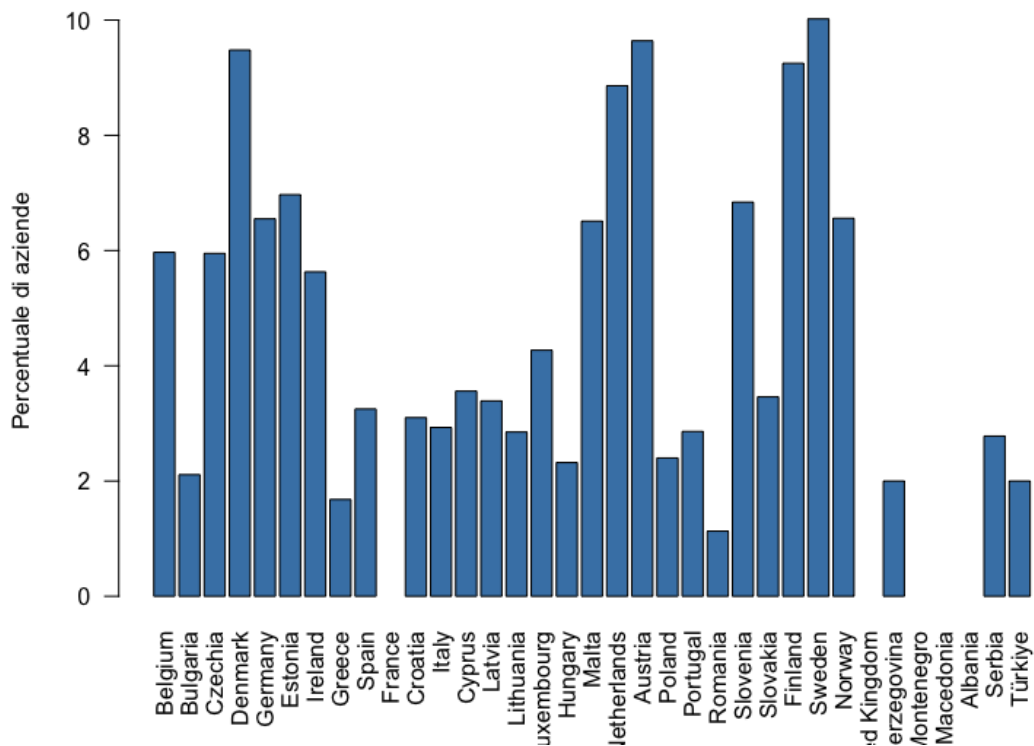
```
barplot(
  height = dataset_E_AI_PMS$'2023',
  names.arg = dataset_E_AI_PMS$GEO_label,
  col = "red",
  main = "Percentuale di aziende che usano l'AI per marketing and sales 2023",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)
```

Percentuale di aziende che usano l'AI per marketing and sales 2023



```
barplot(
  height = dataset_E_AI_PMS$'2024',
  names.arg = dataset_E_AI_PMS$GEO_label,
  col = "steelblue",
  main = "Percentuale di aziende che usano l'AI per marketing and sales 2024",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)
```

Percentuale di aziende che usano l'AI per marketing and sales 2024

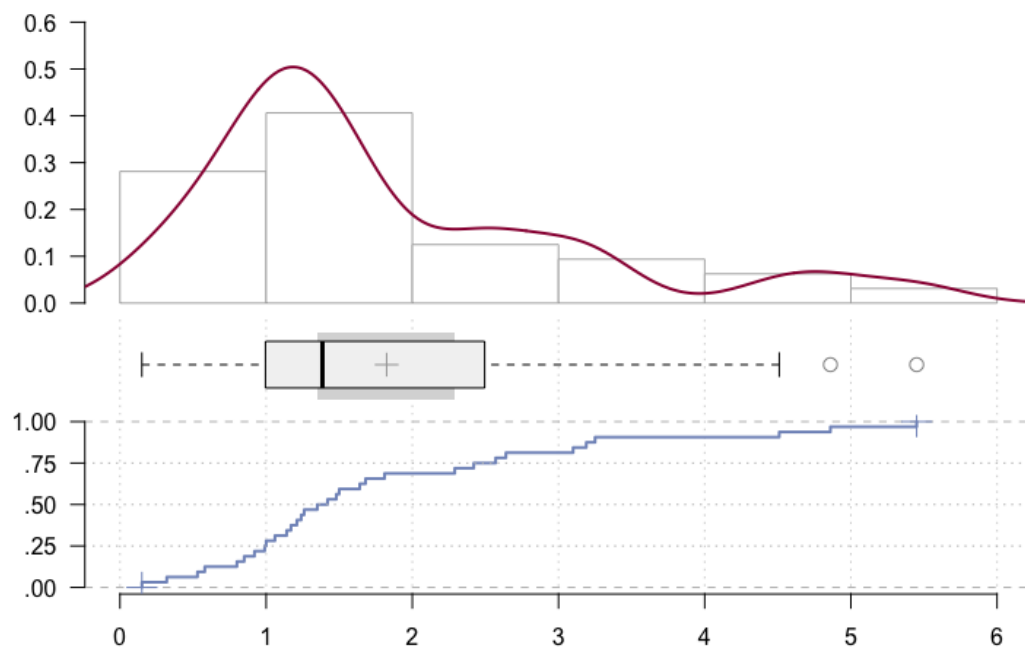


```
Desc(dataset_E_AI_PMS[,2:4])

## -----
## Describe dataset_E_AI_PMS[, 2:4] (tbl_df, tbl, data.frame):
##
## data frame: 35 obs. of 3 variables
##      29 complete cases (82.9%)
##
##   Nr  Class ColName  NAs      Levels
##   1   num   2021     3 (8.6%)
##   2   num   2023     4 (11.4%)
##   3   num   2024     5 (14.3%)
##
## -----
## 1 - 2021 (numeric)
##
##   length      n      NAs  unique      0s      mean  meanCI'
##      35      32       3      = n      0  1.8244  1.3549
##           91.4%   8.6%      0.0%      2.2938
##
##   .05   .10   .25  median   .75   .90   .95
## 0.4355 0.6020 0.9975 1.3850 2.4575 3.2440 4.6675
##
##   range      sd  vcoef      mad      IQR      skew      kurt
## 5.3000 1.3021 0.7137 0.7413 1.4600 1.2146 0.7389
```

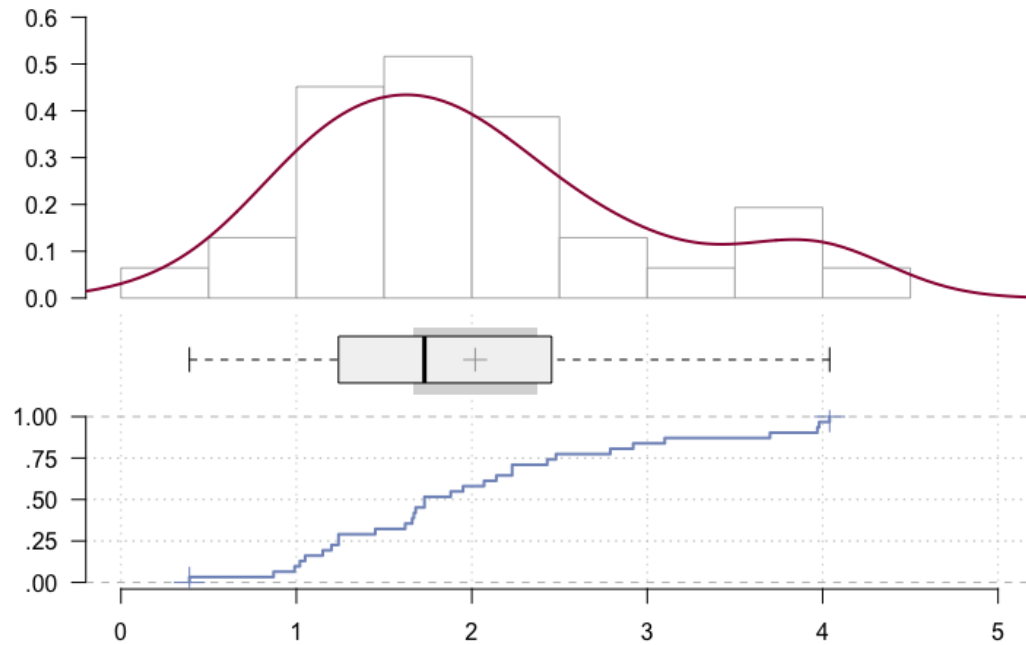
```
##
## lowest : 0.15, 0.32, 0.53, 0.58, 0.8
## highest: 3.19, 3.25, 4.51, 4.86, 5.45
##
## ' 95%-CI (classic)
```

1 - 2021 (numeric)



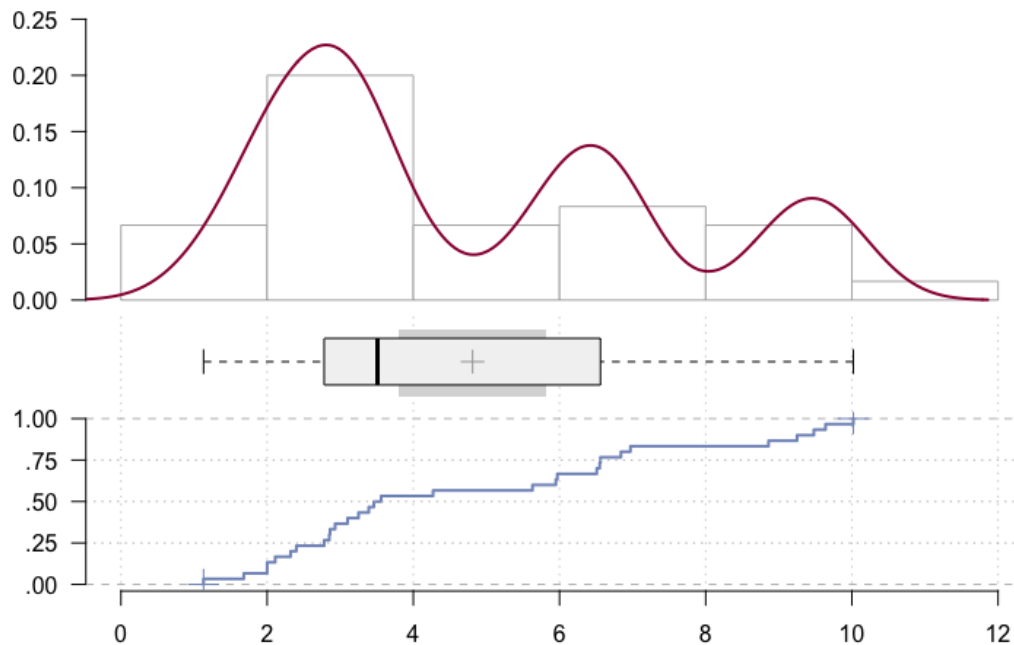
```
##
## 2 - 2023 (numeric)
##
##   length      n    NAs  unique      0s   mean  meanCI '
##      35      31      4      28      0  2.019   1.663
##      88.6%  11.4%      0.0%
##
##   .05   .10   .25  median   .75   .90   .95
##  0.930  1.020  1.240  1.730  2.455  3.700  3.975
##
##   range     sd  vcoef     mad    IQR   skew   kurt
##   3.650  0.971  0.481   0.786  1.215  0.676 -0.454
##
## lowest : 0.39, 0.87, 0.99, 1.02, 1.05
## highest: 3.1, 3.7, 3.97, 3.98, 4.04
##
## ' 95%-CI (classic)
```

2 - 2023 (numeric)



```
##
## 3 - 2024 (numeric)
##
##   length      n    NAs  unique      0s    mean  meanCI '
##      35      30      5      29      0  4.8107  3.7957
##           85.7%  14.3%      0.0%      5.8256
##
##   .05   .10   .25  median   .75   .90   .95
##  1.8240  2.0000  2.7975  3.5100  6.5575  9.2730  9.5680
##
##   range      sd  vcoef      mad      IQR      skew      kurt
##   8.8900  2.7182  0.5650  2.4759  3.7600  0.5425 -1.1097
##
## lowest : 1.13, 1.68, 2.0 (2), 2.11, 2.32
## highest: 8.86, 9.25, 9.48, 9.64, 10.02
##
## ' 95%-CI (classic)
```

3 - 2024 (numeric)



2.5 Analisi Specifica: Imprese che Usano AI per la Logistica

Un ulteriore approfondimento è stato condotto sulle imprese che utilizzano tecnologie di intelligenza artificiale per migliorare la logistica (domanda E_AI_PLOG). L'analisi si basa sui dati disponibili per gli anni 2021, 2023 e 2024 e riguarda tutte le aziende con più di 10 dipendenti. Sono stati generati barplot annuali per evidenziare le variazioni nell'adozione dell'AI per la logistica tra i diversi paesi. Inoltre, è stata fornita una descrizione statistica riassuntiva per esaminare l'evoluzione dell'uso dell'AI in questo ambito e le differenze tra i vari stati nel tempo.

```
dataset_E_AI_PLOG <- read_excel("ISOC_E_AI_PLOG.xlsx")
```

```
dataset_E_AI_PLOG[, c(2, 3,4)] <- lapply(dataset_E_AI_PLOG[, c(2, 3,4)],  
as.numeric)
```

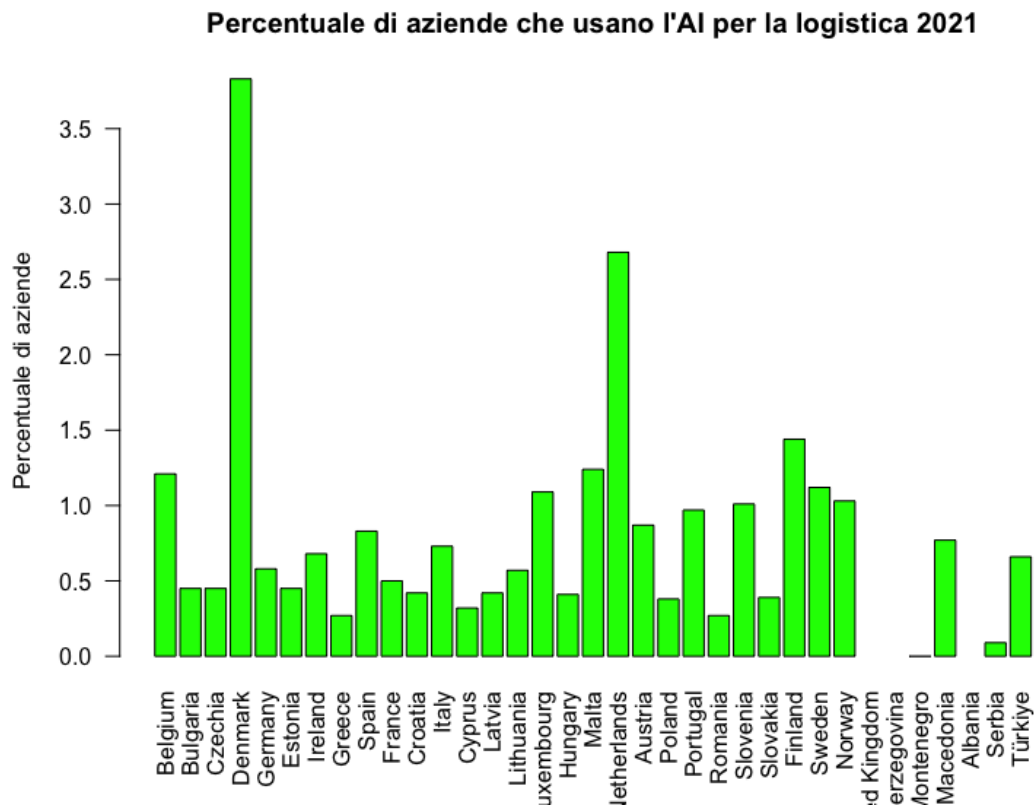
```
## Warning in lapply(dataset_E_AI_PLOG[, c(2, 3, 4)], as.numeric): NAs introduced  
## by coercion  
## Warning in lapply(dataset_E_AI_PLOG[, c(2, 3, 4)], as.numeric): NAs introduced  
## by coercion  
## Warning in lapply(dataset_E_AI_PLOG[, c(2, 3, 4)], as.numeric): NAs introduced  
## by coercion
```

```
dataset_E_AI_PLOG <- dataset_E_AI_PLOG[-1,]  
colnames(dataset_E_AI_PLOG)[1] <- "GEO_label"
```

```

barplot(
  height = dataset_E_AI_PLOG$'2021',
  names.arg = dataset_E_AI_PLOG$GEO_label,
  col = "green",
  main = "Percentuale di aziende che usano l'AI per la logistica 2021",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)

```

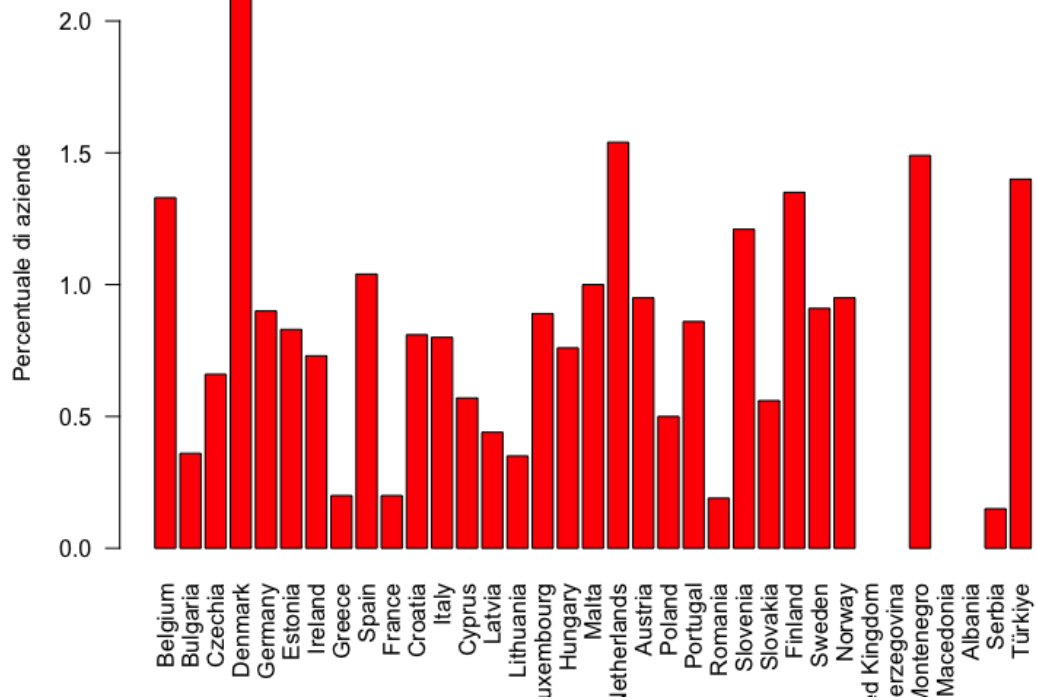


```

barplot(
  height = dataset_E_AI_PLOG$'2023',
  names.arg = dataset_E_AI_PLOG$GEO_label,
  col = "red",
  main = "Percentuale di aziende che usano l'AI per la logistica 2023",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)

```


Percentuale di aziende che usano l'AI per la logistica 2023

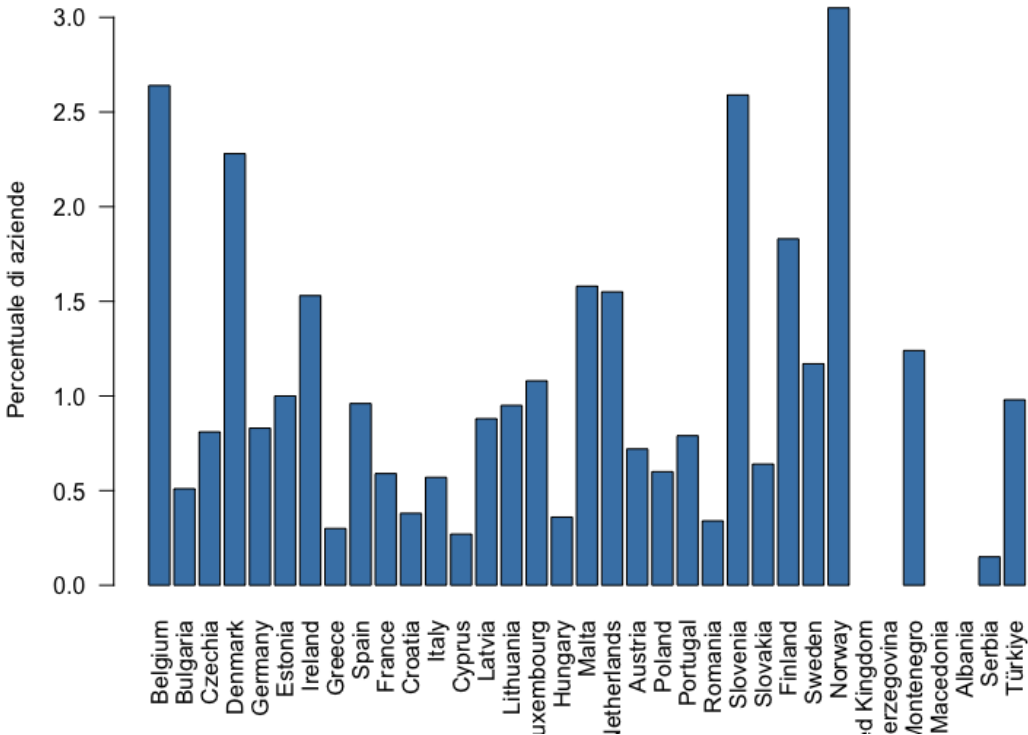


```

barplot(
  height = dataset_E_AI_PLOG$'2024',
  names.arg = dataset_E_AI_PLOG$GEO_label,
  col = "steelblue",
  main = "Percentuale di aziende che usano l'AI per la logistica 2024",
  ylab = "Percentuale di aziende",
  las = 2 # Ruota le etichette sull'asse X
)

```

Percentuale di aziende che usano l'AI per la logistica 2024

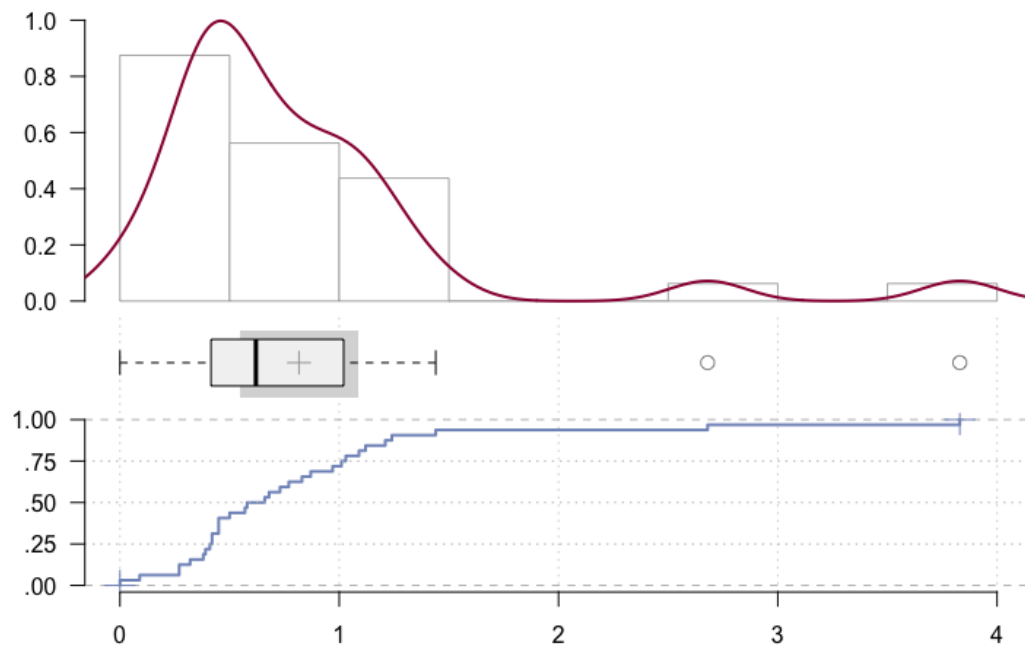


```
Desc(dataset_E_AI_PLOG[,2:4])
```

```
## -----
## Describe dataset_E_AI_PLOG[, 2:4] (tbl_df, tbl, data.frame):
##
## data frame: 35 obs. of 3 variables
## 31 complete cases (88.6%)
##
##   Nr  Class ColName  NAs      Levels
## 1   num   2021      3 (8.6%)
## 2   num   2023      4 (11.4%)
## 3   num   2024      4 (11.4%)
##
## -----
## 1 - 2021 (numeric)
##
##   length      n      NAs  unique      0s      mean  meanCI'
##      35      32        3      28        1  0.8166  0.5487
##           91.4%    8.6%           2.9%           1.0845
##
##   .05   .10   .25  median   .75   .90   .95
## 0.1890 0.2750 0.4175 0.6200 1.0150 1.2370 1.9980
##
##   range      sd  vcoef      mad      IQR      skew      kurt
## 3.8300 0.7431 0.9100 0.3632 0.5975 2.4632 6.8329
```

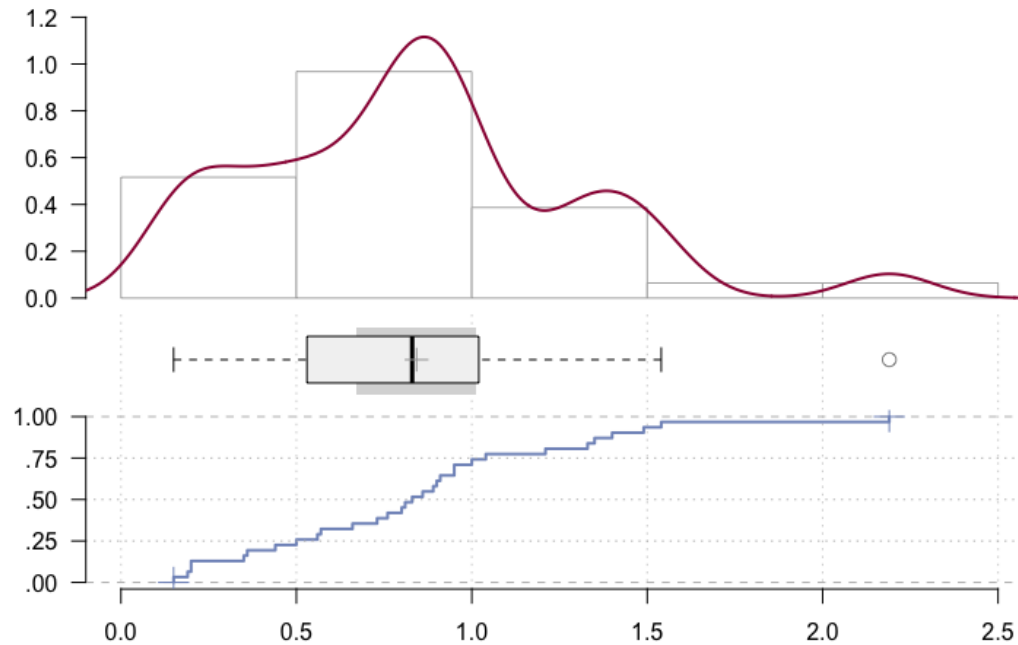
```
##
## lowest : 0.0, 0.09, 0.27 (2), 0.32, 0.38
## highest: 1.21, 1.24, 1.44, 2.68, 3.83
##
## ' 95%-CI (classic)
```

1 - 2021 (numeric)



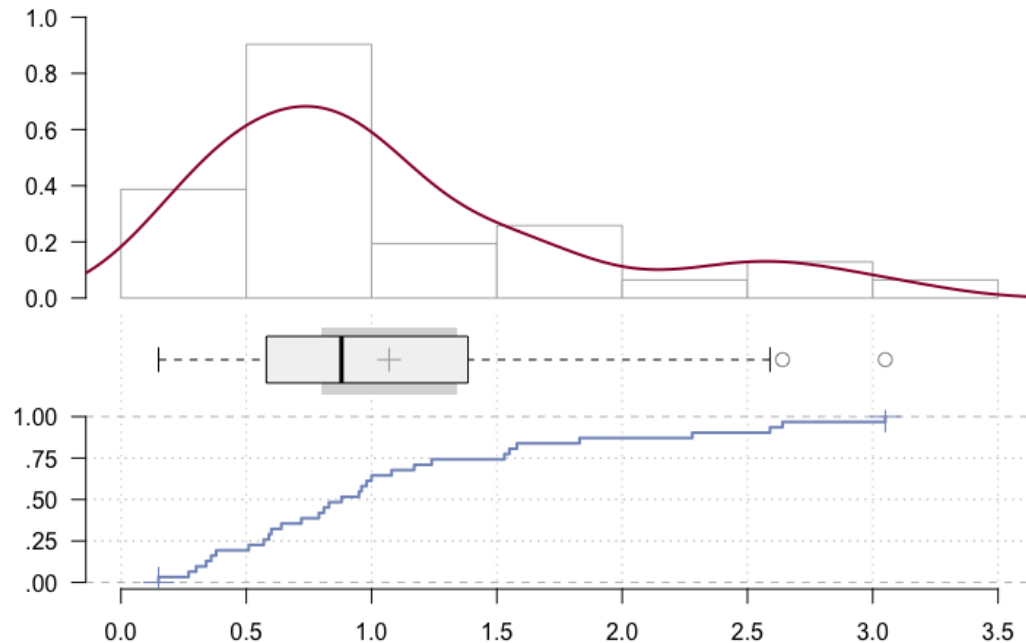
```
##
## 2 - 2023 (numeric)
##
##   length      n    NAs  unique      0s    mean  meanCI '
##      35      31      4      29      0  0.843  0.672
##      88.6%  11.4%      0.0%      1.013
##
##   .05   .10   .25  median   .75   .90   .95
##  0.195 0.200 0.530  0.830  1.020  1.400  1.515
##
##   range      sd  vcoef      mad    IQR    skew    kurt
##  2.040  0.464  0.551   0.400  0.490  0.650  0.391
##
## lowest : 0.15, 0.19, 0.2 (2), 0.35, 0.36
## highest: 1.35, 1.4, 1.49, 1.54, 2.19
##
## ' 95%-CI (classic)
```

2 - 2023 (numeric)



```
##
## 3 - 2024 (numeric)
##
##   length      n    NAs  unique      0s   mean  meanCI '
##      35       31      4    = n      0  1.070  0.797
##          88.6%  11.4%          0.0%      1.343
##
##   .05   .10   .25  median   .75   .90   .95
##  0.285  0.340  0.580  0.880  1.385  2.280  2.615
##
##   range     sd  vcoef     mad    IQR   skew   kurt
##   2.900   0.745  0.697   0.534  0.805  1.104  0.342
##
## lowest : 0.15, 0.27, 0.3, 0.34, 0.36
## highest: 1.83, 2.28, 2.59, 2.64, 3.05
##
## ' 95%-CI (classic)
```

3 - 2024 (numeric)



2.6 Analisi correlazione

Si vuole studiare l'associazione tra il complemento dell'indicatore E_AI_0, cioè l'utilizzo di determinate declinazioni dell'intelligenza artificiale, e la dimensione delle imprese per numero di dipendenti. Per effettuare l'analisi, è stata resa la variabile `size_emp` (rappresentante la classe di numero di dipendenti) un fattore (variabile ordinale) con tre livelli:

- livello 1: dai 10 ai 49 dipendenti
- livello 2: dai 50 ai 249 dipendenti
- livello 3: più di 250 dipendenti

Sulla misura di associazione da calcolare, si è considerato che non è appropriato utilizzare la correlazione di Pearson se una delle variabili è di tipo fattoriale (categorica) e l'altra è numerica. La correlazione di Pearson, infatti, si basa sul concetto di relazione lineare e richiede che entrambe le variabili siano quantitative, cioè misurate su scale numeriche. In questo caso risultava impossibile connotare i valori della variabile fattoriale in termini di distanza tra loro. Inoltre, per valutare la significatività statistica del coefficiente attraverso test parametrici, occorrerebbe che le variabili fossero distribuite normalmente, ciò non può essere verificato dalla variabile di tipo fattoriale. Tuttavia, essendo la variabile fattoriale di tipo ordinale, è possibile utilizzare l'indice di correlazione di Spearman, il quale non impone alcuna assunzione sulla distribuzione delle variabili che si vogliono analizzare.

```

library(readxl)

dataset_E_AI_0_3dipclasses <- read_excel("E_AI_0_3dipclasses.xlsx")

## New names:
## • `` -> `...5`

dataset_E_AI_0_3dipclasses <- dataset_E_AI_0_3dipclasses[, -5]

dataset_E_AI_0_3dipclasses$'size_emp' <-
factor(dataset_E_AI_0_3dipclasses$'size_emp', levels = c("10-49", "50-249",
"GE250"))

#rendere numerica la variabile contenente i valori osservati
dataset_E_AI_0_3dipclasses$'OBS_VALUE' <-
as.numeric(dataset_E_AI_0_3dipclasses$'OBS_VALUE')

#complemento a 100
dataset_E_AI_0_3dipclasses$'%AI' <- 100 - dataset_E_AI_0_3dipclasses$OBS_VALUE

correlazione <- cor(as.numeric(dataset_E_AI_0_3dipclasses$'size_emp'),
                    dataset_E_AI_0_3dipclasses$'%AI',
                    method = "spearman",
                    use = "complete.obs") #trascura righe in cui non sono presenti
dati

# Stampa il risultato

print(correlazione)

## [1] 0.705999

```

Il valore del coefficiente $p=0.705999$, positivo e vicino ad 1, suggerisce la presenza di associazione tra le variabili indagate, in particolare, testimonia che le variabili sono monotonicamente legate (cioè una aumenta costantemente rispetto all'altra, anche in modo non lineare). Dal punto di vista qualitativo ciò significa che le imprese di grandi dimensioni in Europa, utilizzano maggiormente l'intelligenza artificiale; ciò potrebbe essere dovuto alla maggiore facilità con cui possono accedervi o alla presenza in esse di expertise in grado di utilizzare tale tecnologia.

2.7 Analisi in componenti principali

Il dataset utilizzato in quest'analisi deriva da dati presenti sul sito eurostat riguardanti l'utilizzo dell'IA da parte delle aziende. In particolare, l'indicatore analizzato per ciascuno dei paesi UE considerato è "E_AI_TANY" (indica la percentuale di imprese che utilizzano almeno una delle seguenti applicazioni dell'IA: text mining, speech recognition, written text generation, speech generation, identificazione di oggetti basata su immagini, machine learning, automazione dei

processi, movimento fisico delle macchine), i valori all'interno delle celle sono risalenti al 2024 e le variabili rappresentano le attività economiche di riferimento, ossia i settori industriali in cui operano le aziende su cui sono state calcolate le statistiche presenti nelle celle.

Tra le attività economiche sono state eliminate "Electricity&Gas", "Water Supply" e "Real Estate" a causa della mancanza di dati riguardanti una significativa porzione di paesi. Infine, per il settore "Scientific&Technic Activities", non era presente il valore dell'indice per il Portogallo, dunque è stato aggiunto nella cella corrispondente il valore della mediana dei valori della colonna.

```
library(readxl)
df <-
read_excel("C:/Users/1jaco/OneDrive/Desktop/SMA/Progetto_SMA/Dataset_E_AI_TANY2024_
byNACE.xlsx")
library("FactoMineR")
library("factoextra")

## Caricamento del pacchetto richiesto: ggplot2

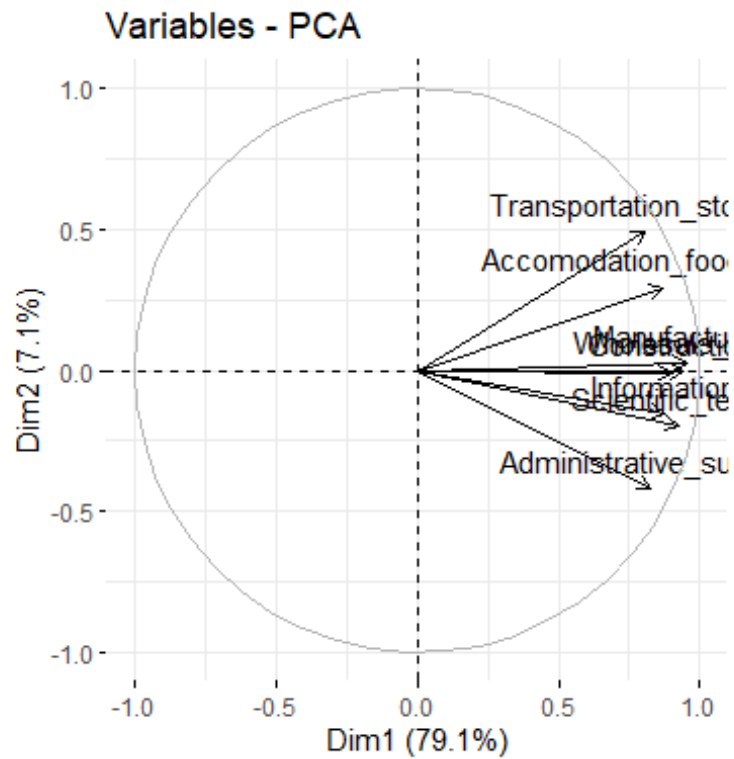
## Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa

res.pca <- PCA(df[,2:9], graph = FALSE)

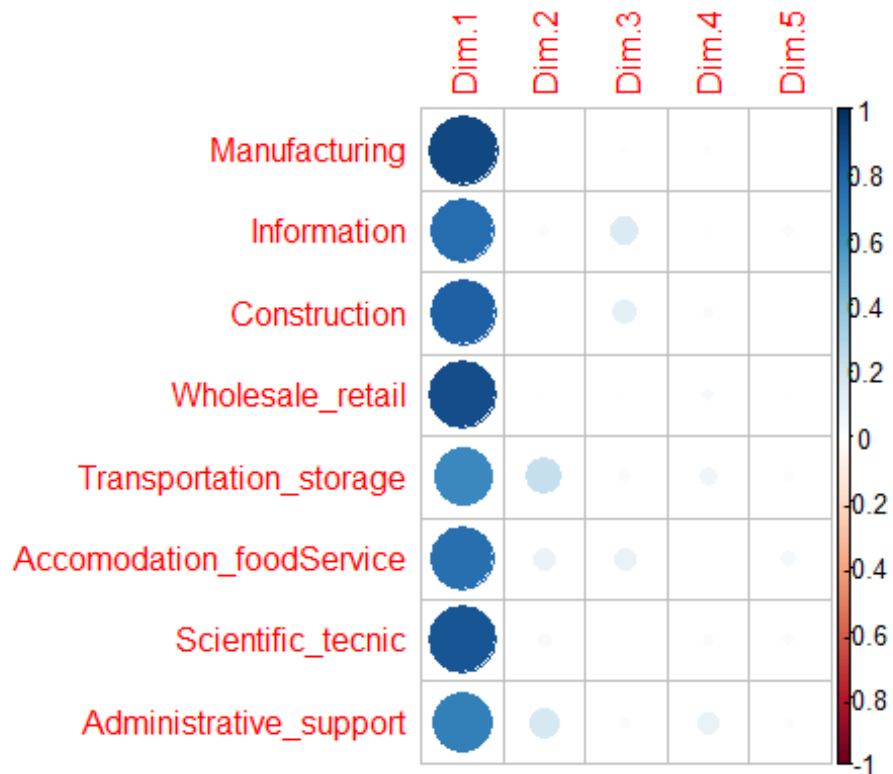
eig.val <- get_eigenvalue(res.pca)
eig.val

##          eigenvalue variance.percent cumulative.variance.percent
## Dim.1 6.32703986      79.0879982      79.08800
## Dim.2 0.56488800       7.0611000      86.14910
## Dim.3 0.41767511       5.2209388      91.37004
## Dim.4 0.25346202       3.1682752      94.53831
## Dim.5 0.14702223       1.8377779      96.37609
## Dim.6 0.12763476       1.5954345      97.97152
## Dim.7 0.09907041       1.2383802      99.20990
## Dim.8 0.06320762       0.7900952     100.00000

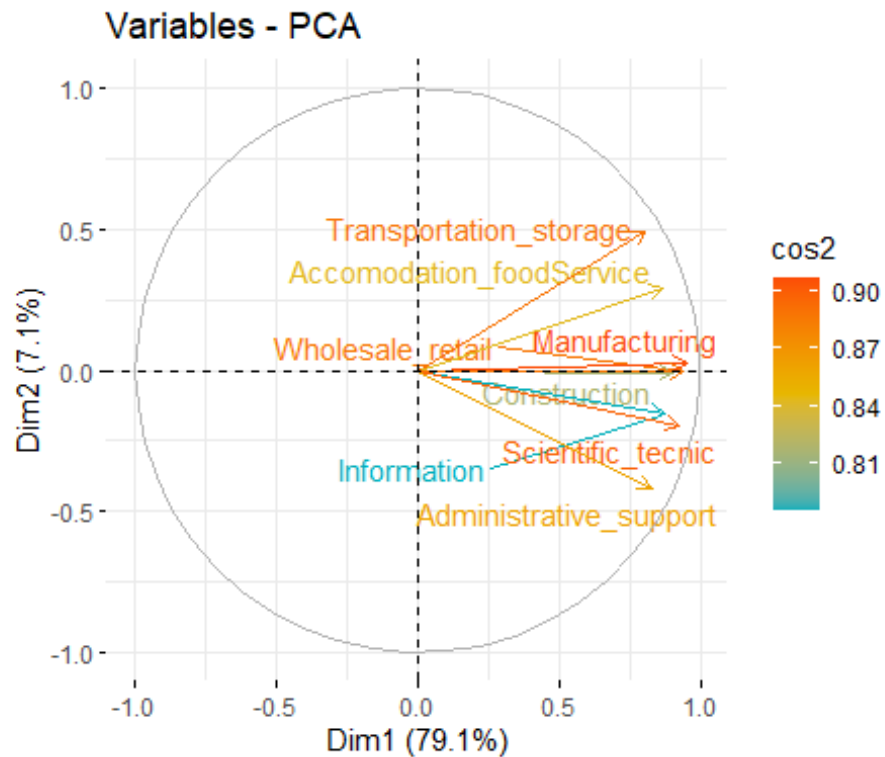
fviz_pca_var(res.pca, col.var = "black")
```



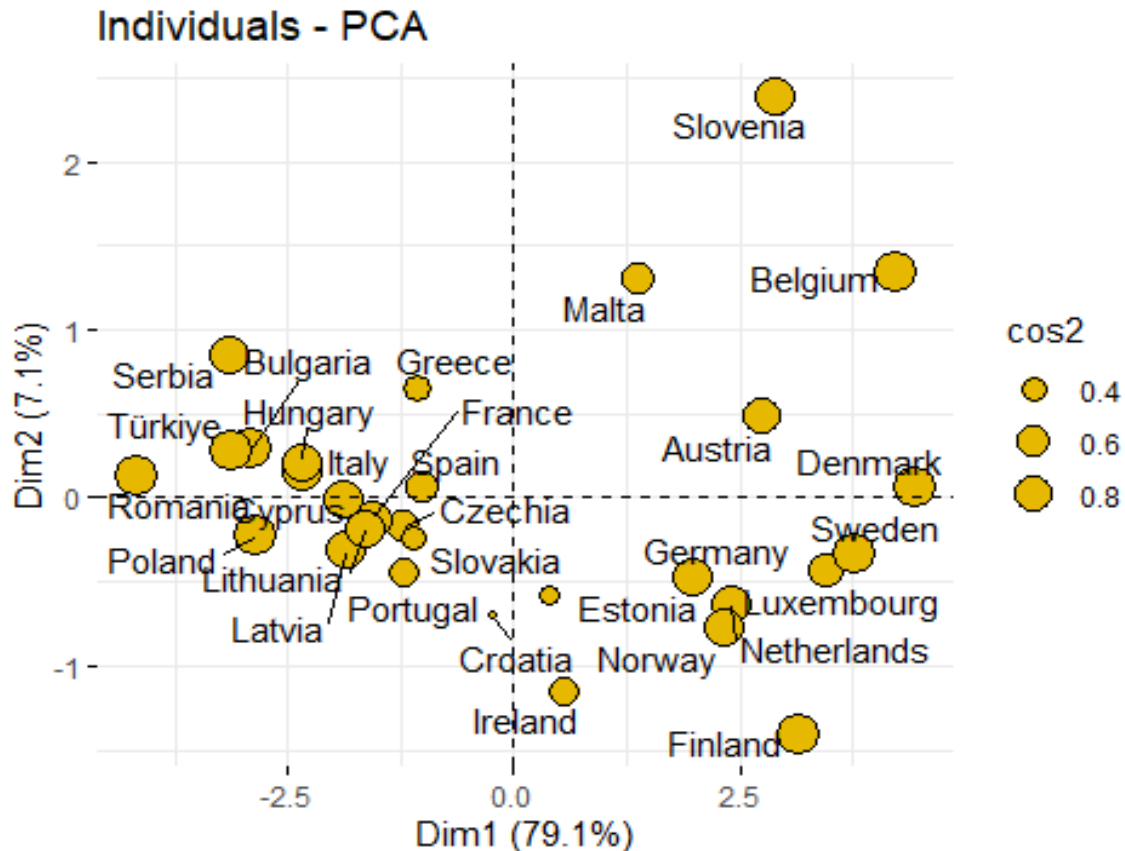
```
library("corrplot")
## corrplot 0.95 loaded
corrplot(res.pca$var$cos2, is.corr=TRUE)
```




```
fviz_pca_var(res.pca, col.var = "cos2", gradient.cols = c("#00AFBB", "#E7B800",
"#FC4E07"), repel = TRUE)
```



```
rownames(res.pca$ind$coord)<-df$Country
fviz_pca_ind(res.pca, pointsize = "cos2",
pointshape = 21 , fill = "#E7B800",
repel = TRUE)
```



Il risultato dell'analisi in componenti principali, a partire dalle otto variabili presenti nel dataset, individua due dimensioni che in totale riescono a spiegare circa il 90% della totale variabilità dei dati. In particolare, la prima dimensione risulta particolarmente esplicativa, infatti nella rappresentazione all'interno di un piano dim1-dim2, i diversi paesi risultano quasi completamente collocati in prossimità dell'asse dim1. Alla formazione di tale dimensione contribuiscono in modo rilevante tutte le variabili del dataset, come è possibile notare dal correlogramma rappresentato.

La rappresentazione grafica del piano con le prime due dimensioni è molto esplicativa in quanto fa emergere che i paesi sono divisi in due principali categorie, tale fatto è supportato soprattutto dai valori della coordinata rappresentante la prima dimensione. E' possibile affermare che per le aziende del gruppo di paesi contraddistinti da maggiori valori della coordinata rappresentata dalla prima dimensione si ha un maggiore sviluppo nell'utilizzo dell'IA, infatti paesi appartenenti a questa categoria, come la Danimarca e l'Olanda hanno una percentuale media di utilizzo dell'IA tra i vari settori maggiore rispetto a quella degli altri paesi.

Dalla collocazione dei paesi all'interno del piano contenente le prime due dimensioni, la Slovenia rappresenta un'eccezione in quanto la sua coordinata della seconda dimensione è significativamente maggiore di 0; ciò testimonia il fatto che le aziende di tale paese hanno una percentuale di utilizzo dell'IA diversa rispetto a quella degli altri paesi dell'UE nei settori che maggiormente contribuiscono alla formazione della seconda dimensione, che dal correlogramma risultano essere "administrative support" e "transportation storage".