Spesa del governo sull'istruzione: può l'istruzione da sola risolvere i problemi economici?

Jozef Jaku 2024-01-30

In questa analisi studieremo se un aumento della spesa pubblica in educazione si traduce in una crescita economica misurabile e in una riduzione dei tassi di disoccupazione

Caricamento delle librerie

```
# Caricamento delle librerie
library(tidyverse)
library(plotly)
```

Caricamento dei dataset

```
# Caricamento della spesa pubblica
spesa_pubblica <- read.csv('govt_edu_expenditure.csv', skip = 4, header = FALSE)</pre>
# Impostazione dei nomi delle colonne alla prima riga
colnames(spesa_pubblica) <- as.character(spesa_pubblica[1, ])</pre>
# Rimozione della prima riga dai dati
spesa_pubblica <- spesa_pubblica[-1, ]</pre>
# Conversione del dataframe in tipo carattere
spesa_pubblica <- as.data.frame(lapply(spesa_pubblica, as.character), stringsAsFactors = FALSE)</pre>
# Pivot del dataframe
spesa_pubblica <- spesa_pubblica %>%
 pivot_longer(cols = -Country.Name,
              names_to = "Year",
              values_to = "Value") %>% drop_na()
# Verifica se il primo carattere della variabile Year è 'X' e conserva solo quelle righe
spesa_pubblica <- subset(spesa_pubblica, substr(Year, 1, 1) == 'X')</pre>
# Rinomina il valore in expenditure e rimuove X da Year
spesa_pubblica <- spesa_pubblica %>%
 rename(expenditure = Value) %>%
 mutate(Year = gsub("X", "", Year))
```

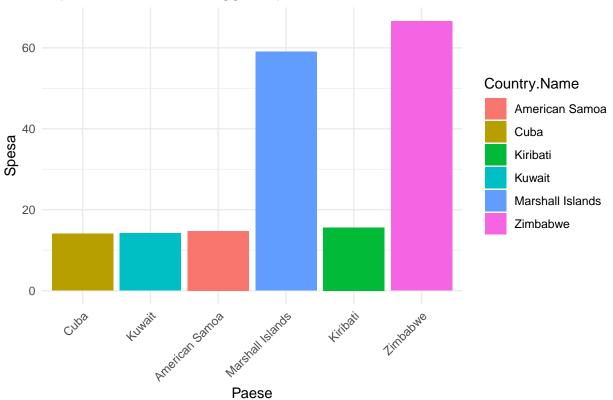
```
# Panoramica dei dati
head(spesa_pubblica, 5)
## # A tibble: 5 x 3
   Country.Name Year expenditure
##
    <chr>
              <chr> <chr>
               1998 4.76316022872925
## 1 Aruba
## 2 Aruba
               1999 4.39526987075806
## 3 Aruba
               2000 4.71536016464233
## 4 Aruba
                2001 4.79850006103516
## 5 Aruba
                2002 4.87177991867065
# Caricamento dei dati sulla crescita economica
crescita_economica <- read.csv('economic_growth.csv', skip = 4, header = FALSE)</pre>
# Impostazione dei nomi delle colonne alla prima riga
colnames(crescita_economica) <- as.character(crescita_economica[1, ])</pre>
# Rimozione della prima riga dai dati
crescita_economica <- crescita_economica[-1, ]</pre>
# Conversione del dataframe in tipo carattere
crescita_economica <- as.data.frame(lapply(crescita_economica, as.character), stringsAsFactors = FALSE)</pre>
# Pivot del dataframe
crescita_economica <- crescita_economica %>%
 pivot_longer(cols = -Country.Name,
             names to = "Year",
             values_to = "Value") %>% drop_na()
# Verifica se il primo carattere della variabile Year è 'X' e conserva solo quelle righe.
crescita_economica <- subset(crescita_economica, substr(Year, 1, 1) == 'X')</pre>
# Rinomina il valore in expenditure e rimuove X da Year
crescita_economica <- crescita_economica %>%
 rename(economic_growth_rate = Value) %>%
 mutate(Year = gsub("X", "", Year))
# Panoramica dei dati
head(crescita_economica, 5)
## # A tibble: 5 x 3
   Country.Name Year economic_growth_rate
    <chr>
          <chr> <chr>
## 1 Aruba
               1987 16.0784313725487
               1988 18.6486486486489
## 2 Aruba
## 3 Aruba
               1989 12.1298405466969
## 4 Aruba
                1990 3.96140172676478
## 5 Aruba
                1991 7.96287174444331
# Caricamento dei dati sull'educazione
```

```
dati_edu <- read.csv('secondary_enrollment_rate.csv')</pre>
# Rimuovi le colonne non necessarie
dati_edu <- dati_edu %>%
 select(-`Code`) %>%
 rename(Country.Name = Entity,
        enrollment_rate = Combined.total.net.enrolment.rate..both.sexes)
# Panoramica dei dati
head(dati_edu, 5)
    Country.Name Year enrollment_rate
## 1 Afghanistan 1993
## 2 Afghanistan 1820
                               0.000
## 3 Afghanistan 1825
                               0.000
## 4 Afghanistan 1830
                               0.000
## 5 Afghanistan 1835
                               0.000
# Caricamento dei dati sulla disoccupazione
dati_disoccupazione <- read.csv('Unemployment.csv')</pre>
# Conversione del dataframe in tipo carattere
dati_disoccupazione <- as.data.frame(lapply(dati_disoccupazione, as.character), stringsAsFactors = FALS
# Pivot del dataframe
dati_disoccupazione <- dati_disoccupazione %>%
 select(-Country.Code) %>%
 pivot longer(cols = -Country.Name,
              names_to = "Year",
              values_to = "Value") %>% drop_na()
# Rinomina il valore in unemployment_rate e rimuove X da Year
dati_disoccupazione <- dati_disoccupazione %>%
 rename(unemployment_rate = Value) %>%
 mutate(Year = gsub("X", "", Year))
# Panoramica dei dati
head(dati_disoccupazione, 5)
## # A tibble: 5 x 3
##
    Country.Name Year unemployment_rate
##
    <chr>
                <chr> <chr>
## 1 Aruba
                2000 6.92
## 2 Aruba
               2001 6.9
## 3 Aruba
                2007 5.71
## 4 Aruba
                2010 10.6
## 5 Aruba
                2011 8.9
```

Paesi con le maggiori spese in educazione (% del PIL)

```
# Top 10 dei paesi con la maggior spesa
spesa_pubblica_top10 <- spesa_pubblica %>%
  # Converti 'expenditure' in numerico
  mutate(expenditure = as.numeric(expenditure)) %>%
  # Organizza in ordine decrescente
  arrange(desc(expenditure)) %>%
  # Seleziona le prime 10 righe
  slice_head(n = 10)
# Creazione di un grafico a barre con i primi 10 paesi
ggplot(spesa_pubblica_top10,
       aes(x = reorder(Country.Name, expenditure),
           y = expenditure,
           fill = Country.Name)) +
  geom_bar(stat = "identity") +
  labs(x = "Paese",
      y = "Spesa",
       title = "Top 10 Paesi con la Maggior Spesa Governativa in Educazione") +
  theme minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Top 10 Paesi con la Maggior Spesa Governativa in Educazione



Il Zimbabwe è il paese che ha speso di più in educazione, investendo fino al 44% e al 22% del proprio PIL in educazione rispettivamente nel 1994 e nel 1992, seguito da Kiribati che ha speso circa il 15,6% del suo PIL nel 2021. Il terzo paese è le Isole Marshall con circa il 15,2% nel 2018.

Andamento nel tempo della spesa pubblica nell'educazione

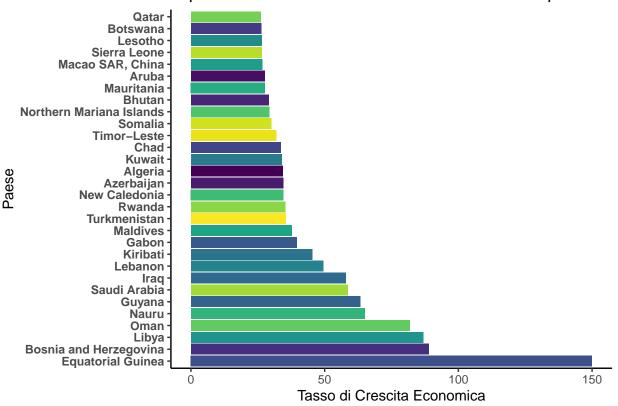
```
# Creazione del grafico delle medie della spesa nel tempo
medie_spesa_pubblica <- spesa_pubblica %>%
  # Assicura che 'Year' sia numerico
  mutate(Year = as.numeric(Year),
         # Converti 'expenditure' in numerico
         expenditure = as.numeric(expenditure)) %>%
  group_by(Year) %>%
  summarize(mean_expenditure = mean(expenditure, na.rm = TRUE))
# Creazione di un grafico a linee
grafico_spesa_pubblica <- medie_spesa_pubblica %>%
  ggplot(aes(x = Year, y = mean_expenditure)) +
  geom_line(size = 1, color = "skyblue") +
  geom_point(size = 2, color = "black") +
  theme minimal() +
  theme(axis.text.x = element_text(face = "bold"),
        axis.text.y = element_text(face = "bold")) +
  ylab("Media della Spesa Governativa")
# Converti ggplot in un grafico interattivo
grafico spesa pubblica int <- grafico spesa pubblica %>%
  ggplotly() %>%
  config(displayModeBar = FALSE) # Nascondi la barra dei comandi per un aspetto più pulito
# Output: grafico interattivo
grafico_spesa_pubblica_int
Media della Spesa Governativa
   4.8
   4.0
       1970
                  1980
                             1990
                                        2000
                                                   2010
                                                              2020
                                   Year
```

La spesa è stata altalenante dal 1970 al 2020, con il 1994 che ha registrato la spesa più elevata e il 2020 che ha registrato la spesa più bassa. Questo potrebbe essere dovuto alla crisi finanziaria del 1994 e agli effetti della pandemia di COVID-19 nel 2020.

Paese con la maggior crescita economica

```
# Creazione del grafico della crescita economica
crescita_economica_top <- crescita_economica %>%
    # Converti in numerico
  mutate(economic growth rate = as.numeric(economic growth rate)) %>%
  group_by(Country.Name) %>%
  summarize(max_growth_rate = max(economic_growth_rate, na.rm = TRUE)) %>%
  arrange(desc(max_growth_rate)) %>%
  # Seleziona i primi 30 paesi con la crescita più alta
  slice head(n = 30)
# Creazione di un grafico a barre orizzontali
ggplot(crescita_economica_top, aes(x = max_growth_rate,
                                y = reorder(Country.Name, -max_growth_rate),
                                fill = Country.Name)) +
  geom_bar(stat = "identity") +
  labs(x = "Tasso di Crescita Economica",
       y = "Paese",
       title = "Top 10 Paesi con il Tasso di Crescita Economica più Alto") +
  theme_classic() +
  theme(axis.text.y = element text(face = "bold"),
        legend.position = "none") +
  scale_fill_viridis_d()
```

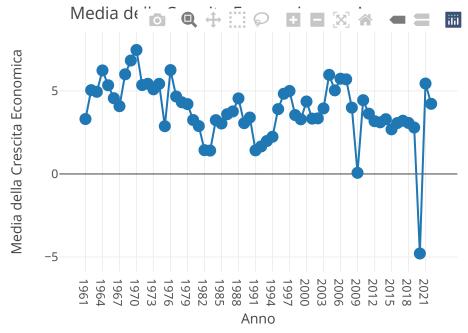
Top 10 Paesi con il Tasso di Crescita Economica più Alto



Il paese che ha registrato la crescita economica più elevata è la Guinea Equatoriale con un record del 150%,

seguito dalla Bosnia ed Erzegovina e dalla Libia in quest'ordine. Dei primi 30 paesi con la crescita economica più elevata, solo alcuni di essi sono nella lista dei maggiori spenditori in educazione, come il Kuwait.

Tendenze della crescita economica nel tempo



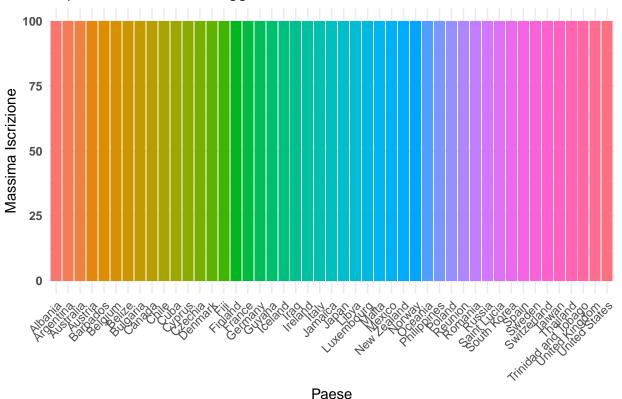
La crescita economica ha oscillato nel tempo e la crescita negativa più elevata si è verificata nel 2020, probabilmente a causa della pandemia di COVID-19

I 30 migliori paesi con il più alto numero di iscrizioni all'istruzione Ad un certo punto

```
# Rimuovi le righe in cui l'iscrizione è 0
dati_edu <- dati_edu %>%
  filter(enrollment_rate != 0)

# Raggruppa per Country.Name e trova l'iscrizione massima per ogni paese
top_paesi <- dati_edu %>%
```

Top 30 Paesi con la Maggior Iscrizione Massima

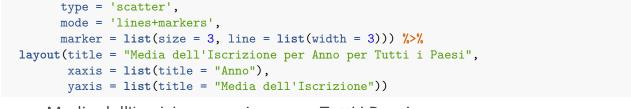


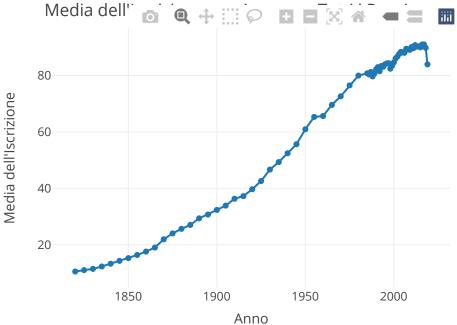
Possiamo vedere che ci sono paesi che hanno avuto iscrizioni del 100% alle scuole secondarie. Questo è piuttosto impressionante.

Tendenza dell'iscrizione nel corso degli anni

```
# Raggruppa per anno e calcola la media dell'iscrizione per ogni anno
iscrizione_per_anno <- dati_edu %>%
  group_by(Year) %>%
  summarize(mean_enrollment = mean(enrollment_rate))

# Crea un grafico a linee interattivo con plotly
plot_ly(iscrizione_per_anno, x = ~Year, y = ~mean_enrollment,
```





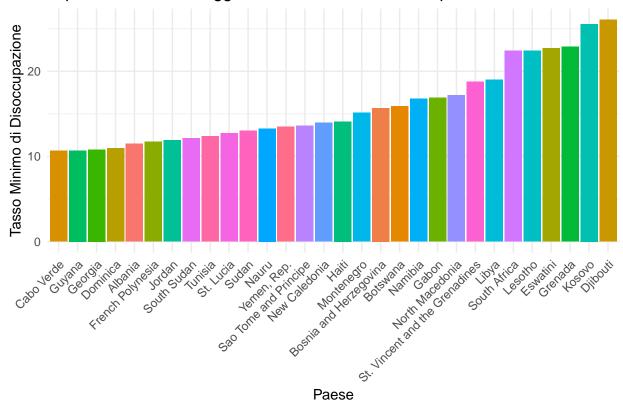
I tassi di iscrizione all'istruzione secondaria sono aumentati nel tempo.

Paesi che hanno avuto la peggiore disoccupazione in un certo momento della storia

```
# Converti in numerico
dati_disoccupazione$Year <- as.numeric(dati_disoccupazione$Year)</pre>
dati_disoccupazione unemployment_rate <- as.numeric(dati_disoccupazione unemployment_rate)
# Raggruppa per Country.Name e trova il tasso di disoccupazione minimo per ogni paese
peggiori_paesi_disoccupazione <- dati_disoccupazione %>%
  group_by(Country.Name) %>%
  summarize(min_unemployment_rate = min(unemployment_rate)) %>%
  # Ordina in ordine decrescente e seleziona i primi 30
  arrange(desc(min_unemployment_rate)) %>%
  slice head(n = 30)
# Crea un grafico a barre orizzontali
ggplot(peggiori_paesi_disoccupazione,
       aes(x = reorder(Country.Name, min_unemployment_rate),
           y = min_unemployment_rate,
           fill = Country.Name)) +
  geom_bar(stat = "identity") +
```

```
labs(x = "Paese",
    y = "Tasso Minimo di Disoccupazione",
    title = "Top 30 Paesi con i Peggiori Tassi Minimi di Disoccupazione") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "none")
```

Top 30 Paesi con i Peggiori Tassi Minimi di Disoccupazione

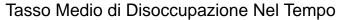


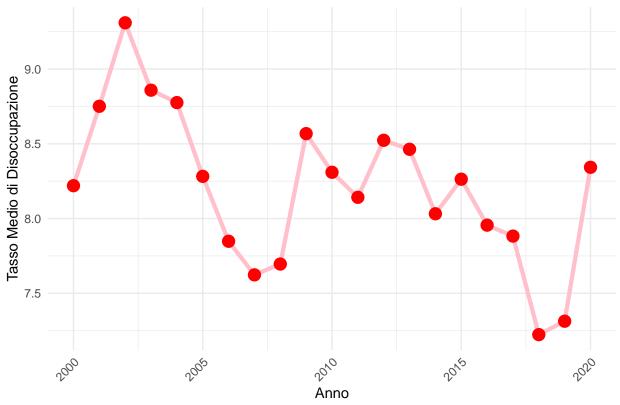
I paesi che hanno avuto la peggiore disoccupazione in un certo momento sono Gibuti, col 26% di disoccupazione, il Kosovo con circa il 25,5% e Grenada con circa il 23%.

Il tasso di disoccupazione sta salendo o scendendo?

```
# Raggruppa per anno e calcola la media del tasso di disoccupazione per ogni anno
media_disoccupazione_per_anno <- dati_disoccupazione %>%
  group_by(Year) %>%
  summarize(mean_unemployment_rate = mean(unemployment_rate, na.rm = TRUE))

# Crea un grafico a linee per la media del tasso di disoccupazione per anno
ggplot(media_disoccupazione_per_anno, aes(x = Year, y = mean_unemployment_rate)) +
  geom_line(size = 1.5, color = "pink") +
  geom_point(size = 4, color = "red") +
  labs(x = "Anno",
        y = "Tasso Medio di Disoccupazione",
        title = "Tasso Medio di Disoccupazione Nel Tempo") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```





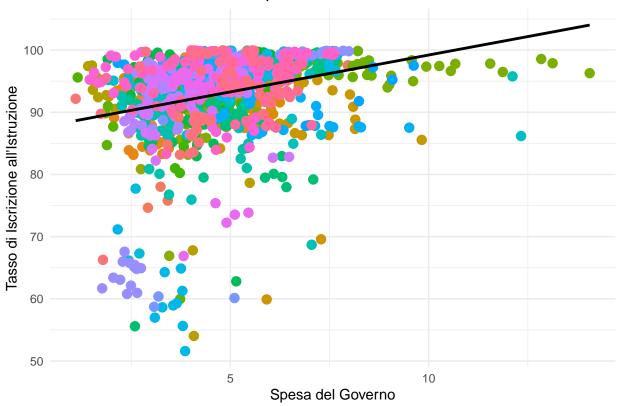
I tassi di disoccupazione sono stati altalenanti nel tempo con alcuni aumenti notevoli intorno al 2008, quando c'è stata una crisi finanziaria, e nel 2020, quando è scoppiata la pandemia.

Visualizzazione dell'ipotesi 1: Correlazione tra la spesa del governo e l'iscrizione all'istruzione

```
# Unisci i dataset su Country.Name e Year
dati_fusione <- merge(spesa_pubblica, crescita_economica, by = c("Country.Name", "Year"))
dati_fusione <- merge(dati_fusione, dati_disoccupazione, by = c("Country.Name", "Year"))
dati_fusione <- merge(dati_fusione, dati_edu, by = c("Country.Name", "Year"))</pre>
# Converti 'enrollment_rate' e 'economic_growth_rate' in numerico
dati_fusione$Year <- as.numeric(dati_fusione$Year)</pre>
dati_fusione$expenditure <- as.numeric(dati_fusione$expenditure)</pre>
dati_fusione$economic_growth_rate <- as.numeric(dati_fusione$economic_growth_rate)
dati_fusione$unemployment_rate <- as.numeric(dati_fusione$unemployment_rate)</pre>
dati_fusione$enrollment_rate <- as.numeric(dati_fusione$enrollment_rate)</pre>
dati_fusione <- dati_fusione[dati_fusione$enrollment_rate >= 50, ]
# Crea un grafico a dispersione con la linea di regressione lineare
ggplot(dati_fusione, aes(x = expenditure,
                        y = enrollment_rate,
                         color = Country.Name)) +
  geom point(size = 3) +
  geom_smooth(method = "lm", se = FALSE, linetype = "solid", color = "black") +
 labs(x = "Spesa del Governo",
```

```
y = "Tasso di Iscrizione all'Istruzione",
    title = "Iscrizione all'Istruzione Vs Spesa del Governo") +
theme_minimal() +
theme(legend.position = "none")
```

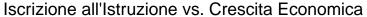
Iscrizione all'Istruzione Vs Spesa del Governo

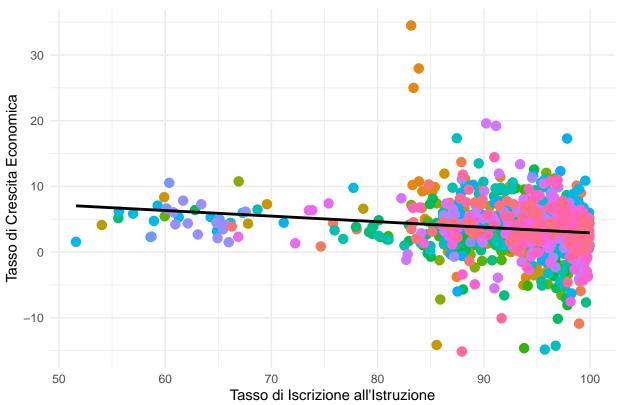


Dall'output sopra, osserviamo che più il governo spende in educazione, più aumentano i tassi di iscrizione. Questo mostra che c'è una correlazione positiva tra la spesa governativa e l'iscrizione.

Quindi quando il governo spende in educazione, l'educazione risolve le difficoltà economiche? Scopriamolo nelle sezioni successive.

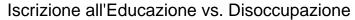
Visualizzazione dell'ipotesi 2: Correlazione tra l'iscrizione all'istruzione e la crescita economica

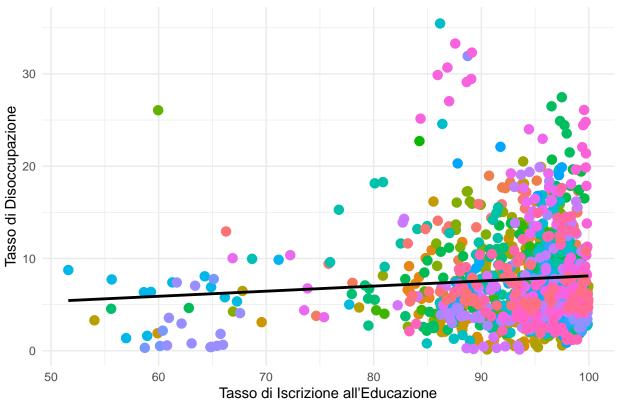




Sorprendentemente, l'analisi rivela una correlazione negativa inaspettata tra i tassi di iscrizione all'istruzione e la crescita economica. Nel contesto dello studio, la correlazione negativa suggerisce che, in media, all'aumentare dei tassi di iscrizione all'istruzione, si verifica una leggera diminuzione della crescita economica. Questa scoperta inaspettata sfida le ipotesi convenzionali secondo cui una maggiore iscrizione all'istruzione dovrebbe portare a una crescita economica più elevata. Potrebbe esserci la necessità di esplorare ulteriormente questa relazione apparentemente contraria alle aspettative.

Visualizzazione dell'Ipotesi 3: Correlazione tra l'Iscrizione all'Educazione e la Disoccupazione





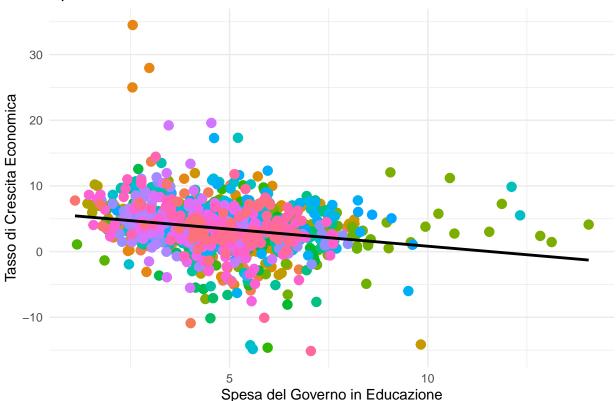
Ancora una volta veniamo sorpresi, l'analisi infatti indica una correlazione positiva tra i tassi di iscrizione all'istruzione secondaria e il tasso di disoccupazione. Esaminando i dati, la relazione positiva significa che all'aumentare dei tassi di iscrizione all'istruzione secondaria, aumenta anche il tasso di disoccupazione. Questa relazione è inaspettata e sfida la comune convinzione che un aumento dell'iscrizione all'istruzione porti automaticamente a tassi di disoccupazione più bassi.

Ma riflettendoci, possiamo notare che quando aumentiamo il numero di lavoratori qualificati senza far crescere l'economia, la maggior parte di loro non troverà lavoro. Ciò significa che più iscrizioni corrispondono a tassi di disoccupazione più elevati.

Visualizzazione dell'Ipotesi 4: Correlazione tra la Spesa in Educazione e la Crescita Economica

```
labs(x = "Spesa del Governo in Educazione",
    y = "Tasso di Crescita Economica",
    title = "Spesa in Educazione vs. Crescita Economica") +
theme_minimal() +
theme(legend.position = "none")
```

Spesa in Educazione vs. Crescita Economica



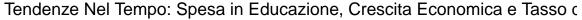
L'analisi rivela una correlazione negativa inaspettata tra la spesa del governo in educazione e il tasso di crescita economica. La correlazione negativa suggerisce che, in media, all'aumentare dell'allocazione di fondi per l'educazione, c'è una tendenza al calo del tasso di crescita economica. Questa relazione sfida la comune convinzione che un maggiore investimento governativo nell'istruzione si traduca direttamente in una crescita economica robusta.

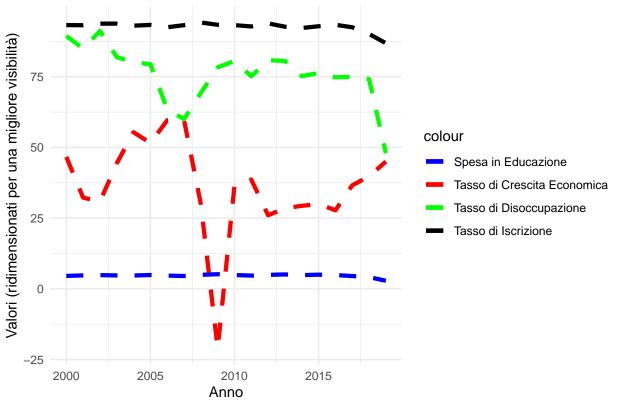
Visualizzazione dell'Ipotesi 5: Spesa in Educazione Nel Tempo e Indicatori Economici

Abbiamo visualizzato un grafico a linee che mostra le tendenze nella spesa del governo in educazione nel tempo, sovrapposte alle tendenze della crescita economica e del tasso di disoccupazione.

Metodo 1

```
media_tasso_disoccupazione = mean(unemployment_rate, na.rm = TRUE),
            media_tasso_iscrizione = mean(enrollment_rate, na.rm = TRUE))
# Crea un grafico a linee con linee multiple
ggplot(dati_raggruppati, aes(x = as.numeric(Year))) +
  geom_line(aes(y = media_spesa,
                color = "Spesa in Educazione"),
            size = 1.5,
            linetype = "dashed") +
  geom_line(aes(y = media_tasso_crescita_economica * 10,
               color = "Tasso di Crescita Economica"),
            size = 1.5, linetype = "dashed") +
  geom_line(aes(y = media_tasso_disoccupazione * 10,
                color = "Tasso di Disoccupazione"),
            size = 1.5,
           linetype = "dashed") +
  geom_line(aes(y = media_tasso_iscrizione,
              color = "Tasso di Iscrizione"),
          size = 1.5,
          linetype = "dashed") +
 labs(x = "Anno",
      y = "Valori (ridimensionati per una migliore visibilità)",
      title = "Tendenze Nel Tempo: Spesa in Educazione, Crescita Economica e Tasso di Disoccupazione")
  scale_color_manual(values = c("Spesa in Educazione" = "blue",
                                "Tasso di Crescita Economica" = "red",
                                "Tasso di Disoccupazione" = "green",
                                "Tasso di Iscrizione" = "black")) +
  theme_minimal()
```

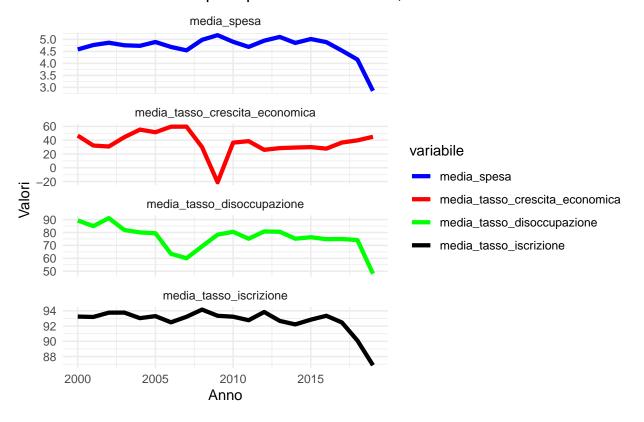




Metodo 2

```
# Raggruppa per Anno e calcola la media di ciascuna variabile
dati_raggruppati <- dati_unificati %>%
  group_by(Year) %>%
  summarize(media_spesa = mean(expenditure, na.rm = TRUE),
            media_tasso_crescita_economica = mean(economic_growth_rate, na.rm = TRUE),
            media_tasso_disoccupazione = mean(unemployment_rate, na.rm = TRUE),
            media_tasso_iscrizione = mean(enrollment_rate, na.rm = TRUE))
# Ridimensiona le variabili per una migliore visibilità
dati_raggruppati$media_tasso_crescita_economica <- dati_raggruppati$media_tasso_crescita_economica * 10
dati_raggruppati$media_tasso_disoccupazione <- dati_raggruppati$media_tasso_disoccupazione * 10
# Ristruttura i dati per ggplot
dati_raggruppati_lungo <- tidyr::gather(dati_raggruppati, key = "variabile", value = "valore", -Year)</pre>
# Crea un grafico a linee con facet_wrap
ggplot(dati_raggruppati_lungo, aes(x = Year,
                              y = valore,
                              color = variabile,
                              linetype = variabile)) + geom_line(size = 1.5) +
  labs(x = "Anno", y = "Valori",
       title = "Tendenze Nel Tempo: Spesa in Educazione, Crescita Economica e Tasso di Disoccupazione")
  scale_color_manual(values = c("media_spesa" = "blue",
                                "media_tasso_crescita_economica" = "red",
```

Tendenze Nel Tempo: Spesa in Educazione, Crescita Economica e Tasso c



Sources

https://data.oecd.org/students/tertiary-graduation-rate.htm#indicator-chart

https://ourworldindata.org/global-education

 ${\rm https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG}$

https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS