

# Stagionalità delle vendite del negozio di strumenti musicali “Toni&Suoni”

Università degli studi di Udine  
Facoltà di IBML  
Corso Fondamenti di Scienza dei Dati e Laboratorio [MA0682]

Studente [MAT159263]  
Fabio Massimo Lattanzio  
Anno accademico 2022- 2023

## INTRODUZIONE

- In questo progetto verrà sviluppata l'**analisi dei dati di vendita**, di un piccolo negozio di strumenti musicali, situato all'interno di una scuola di musica. Il negozio nasce alla fine del 2018 e, oltre alla vendita di strumenti musicali e di tutti i relativi accessori, vende libri musicali (metodi, partiture, canzonieri, quaderni di musica, etc.) e fornisce servizi di liuteria (cambio corde, pulizia dello strumento, accordatura, assemblaggio, sostituzione di meccaniche, etc.).
- Tutti i dati a disposizione, sia relativi al magazzino sia relativi alle vendite del negozio, sono raccolti in un software, utilizzato per la gestione del negozio
- Il mio progetto si focalizzerà sull'**andamento delle vendite** negli anni di osservazione 2019, 2020, 2021, 2022 fino a maggio 2023, con focus su:
  - Andamento delle vendite per anno con particolare attenzione a periodi economici contingenti (covid e chiusure forzate)
  - Stagionalità delle vendite per periodo quali:
    1. apertura anno scolastico;
    2. festività (Natale e Pasqua);
    3. Periodo estivo e chiusura dell'attività scolastica
  - Andamento delle vendite distinta per CATEGORIA (serie: Accessori, Audio, Libri, Liuteria, Strumenti musicali)

## ANALISI DEL DATASET

I dataset estrapolati dal programma di gestione del magazzino MaestroGold sono 2:

1. Magazzino (Movimenti da gennaio 2019 a maggio 2023)
2. Fatturato (Vendite da gennaio 2019 a maggio 2023)

Ho importato i due file .xls in R e li ho salvati nella cartella E:\Data Science\Progetto\Data

I risultati sono salvati nella cartella E:\Data Science\Progetto\Source

Utilizzo il comando `str(Dataset)` che restituisce informazioni sulla struttura dei dataset.

`str(Magazzino)`

```
tibble          [2,035      ×          9]      (S3:      tbl_df/tbl/data.frame)
 $ Codice       : chr [1:2035] "AROMA00AG15A" "EWAVE00GC10" "0278" "TCL019271" ...
 $ Descrizione  : chr [1:2035] "AMPLIFICATORE AROMA PER ACUSTICA AG15 15 WATT" "AMPLIFICATORE CHIT.
ELETTRICA EWAVE00GC10" "BLACKSTAR MINI AMPLI 3W CON ALIMENTATORE" "GUITAR EXAM INITIAL 2020 2023 TRINITY
COLLEGE LONDON"
 $ Categoria    : chr [1:2035] "AMPLIFICATORI" "AMPLIFICATORI" "AMPLIFICATORI" "LIBRI" ...
 $ Ultimo Prezzo : num [1:2035] 81.15 53.28 73.77 9.62 12.3 ...
 $ Quantità     : num [1:2035] 2 2 2 7 12 1 3 30 1 1 ...
 $ Totale Importo: num [1:2035] 162.3 114.8 130.3 65.4 147.5 ...
 $ Utile        : num [1:2035] 61.7 58.2 86.3 29.2 119.5 ...
 $ ult. Carico  : chr [1:2035] "20/07/2020" "20/07/2020" "22/08/2018" "16/09/2020" ...
 $ ult. Scarico : chr [1:2035] "11/12/2020" "17/10/2020" "25/10/2022" "15/09/2022" ...
```

Il dataset Magazzino contiene il numero di articoli in giacenza con indicazione della data ultimo carico, scarico, ultimo prezzo di vendita e relativo utile. Le dimensioni del dataset Magazzino sono: 2035 righe (osservazioni) e 9 colonne (variabili). Non ci sono valori nulli.

## ANALISI DEL DATASET

```
str(Fatturato)
```

```
tibble [8,819 × 12] (S3: tbl_df/tbl/data.frame)
 $ Codice      : chr [1:8819] "B147B" "B692B" "0157" "WP42" ...
 $ Barcode     : chr [1:8819] "8000000109576" "8000000112620" "9002761038798" "9780849750427" ...
 $ Data        : POSIXct[1:8819], format: "2019-01-02" "2019-01-02" "2019-01-02" "2019-01-02" ...
 $ Anno        : num [1:8819] 2019 2019 2019 2019 2019 ...
 $ Mese        : num [1:8819] 1 1 1 1 1 1 1 1 1 ...
 $ Mese_Anno   : num [1:8819] 1 1 1 1 1 1 1 1 1 ...
 $ Categoria   : chr [1:8819] "ACCESSORI" "ACCESSORI" "AUDIO" "LIBRI" ...
 $ Categoria1  : chr [1:8819] "ACCESSORI" "ACCESSORI" "AUDIO" "LIBRI" ...
 $ Descrizione: chr [1:8819] "BACCHETTE 5A" "BACCHETTE 5A" "CUFFIA" "METODO PIANOFORTE" ...
 $ Q.tà        : num [1:8819] 1 1 1 1 1 1 1 1 1 ...
 $ Prezzo      : num [1:8819] 15 2 55 22 12 98 13 3 110 5 ...
 $ Importo     : num [1:8819] 15 2 55 22 12 98 13 3 110 5 ...
```

Il dataset Fatturato contiene pezzi e importi degli articoli venduti giornalmente distinti per categoria.

Le dimensioni del dataset sono: 8819 righe (osservazioni) e 12 colonne (variabili). Non ci sono valori nulli.

## ANALISI DEL DATASET

### Analisi descrittiva univariata

Utilizzo la funzione `summary(Dataset)` per vedere la struttura dei miei 2 dataset.

```
##{r}
summary(Magazzino)
##{r}
```

Codice	Descrizione	Categoria	Ultimo Prezzo	Quantità	Totale Importo
Length:2035	Length:2035	Length:2035	Min. : 0.000	Min. : 1.000	Min. : 0.00
Class :character	Class :character	Class :character	1st Qu.: 9.836	1st Qu.: 1.000	1st Qu.: 19.67
Mode :character	Mode :character	Mode :character	Median : 16.393	Median : 2.000	Median : 45.08
			Mean : 43.043	Mean : 5.691	Mean : 121.25
			3rd Qu.: 41.165	3rd Qu.: 4.000	3rd Qu.: 110.66
			Max. : 860.656	Max. : 729.000	Max. : 12309.83

Utile	Ult. Carico	Ult. Scarico
Min. : -41.033	Length:2035	Length:2035
1st Qu.: 9.711	Class :character	Class :character
Median : 23.052	Mode :character	Mode :character
Mean : 61.711		
3rd Qu.: 56.053		
Max. : 12309.835		

```
##{r}
summary(Fatturato)
##{r}
```

Codice	Barcode	Data	Anno	Mese	Mese_Anno
Length:8819	Length:8819	Min. :2019-01-02 00:00:00.00	Min. :2019	Min. : 1.000	Min. : 1.00
Class :character	Class :character	1st Qu.:2019-12-02 00:00:00.00	1st Qu.:2019	1st Qu.: 3.000	1st Qu.:12.00
Mode :character	Mode :character	Median :2021-02-02 00:00:00.00	Median :2021	Median : 6.000	Median :26.00
		Mean :2021-02-13 06:23:23.42	Mean :2021	Mean : 6.549	Mean :25.98
		3rd Qu.:2022-03-24 00:00:00.00	3rd Qu.:2022	3rd Qu.:10.000	3rd Qu.:39.00
		Max. :2023-05-20 00:00:00.00	Max. :2023	Max. :12.000	Max. :53.00

Categoria	Categoria1	Descrizione	Q.tà	Prezzo	Importo
Length:8819	Length:8819	Length:8819	Min. : 1.000	Min. : 0.00	Min. : 0.00
Class :character	Class :character	Class :character	1st Qu.: 1.000	1st Qu.: 10.00	1st Qu.: 10.00
Mode :character	Mode :character	Mode :character	Median : 1.000	Median : 15.00	Median : 15.00
			Mean : 1.267	Mean : 29.91	Mean : 31.38
			3rd Qu.: 1.000	3rd Qu.: 25.00	3rd Qu.: 25.00
			Max. : 35.000	Max. :1050.00	Max. :1050.00

## ANALISI DEL DATASET

### Analisi descrittiva univariata

Conduco le analisi descrittive delle variabili contenute nel dataset: le variabili Q.tà, Prezzo e Importo sono variabili quantitative, potrò quindi calcolare gli indici di posizione.

```
> summary(Fatturato$Q.tà)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  1.000   1.000   1.267  1.000  35.000
```

La variabile Q.tà assume come valore minimo 1 e come valore massimo 35; la mediana è 1 e la media è 1,267. Sembrerebbe che per la maggior parte delle vendite è richiesto 1 solo articolo, con qualche eccezione.

```
> summary(Fatturato$Prezzo)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00  10.00   15.00   29.91  25.00 1050.00
```

La variabile Prezzo (espressa in Euro) assume come valore minimo 0 e come valore massimo 1050,00€; la media è 29,91€ e la mediana è 15,00€.

```
> summary(Fatturato$Importo)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00  10.00   15.00   31.38  25.00 1050.00
```

La variabile Importo (espressa in Euro) assume come valore minimo 0 e come valore massimo 1050,00€; la media è 31,38€ e la mediana è 15,00€. La differenza tra la variabile Prezzo e la variabile Importo è dovuta al fatto che la prima descrive il prezzo di ogni singolo articolo, la seconda moltiplica il numero degli articoli venduti per il loro prezzo di vendita.

Sia la variabile Prezzo, che la variabile Importo, assumo come minimo il valore 0.

Il fenomeno è dovuto a come sono riportati i dati nel dataset Fatturato che presenta il “mese 16” corrispondente ad aprile 2020, con le variabili Prezzo e Fatturato pari a 0. Si tratta del mese di chiusura totale del negozio per le restrizioni dovute al Covid. Nelle successive analisi sulla stagionalità delle vendite verificherò se la chiusura del negozio ha influito negativamente sul fatturato totale.

## ANALISI DEL DATASET

### Analisi descrittiva univariata

Osservo che media e mediana non coincidono nei vari summary quindi realizzo dei boxplot

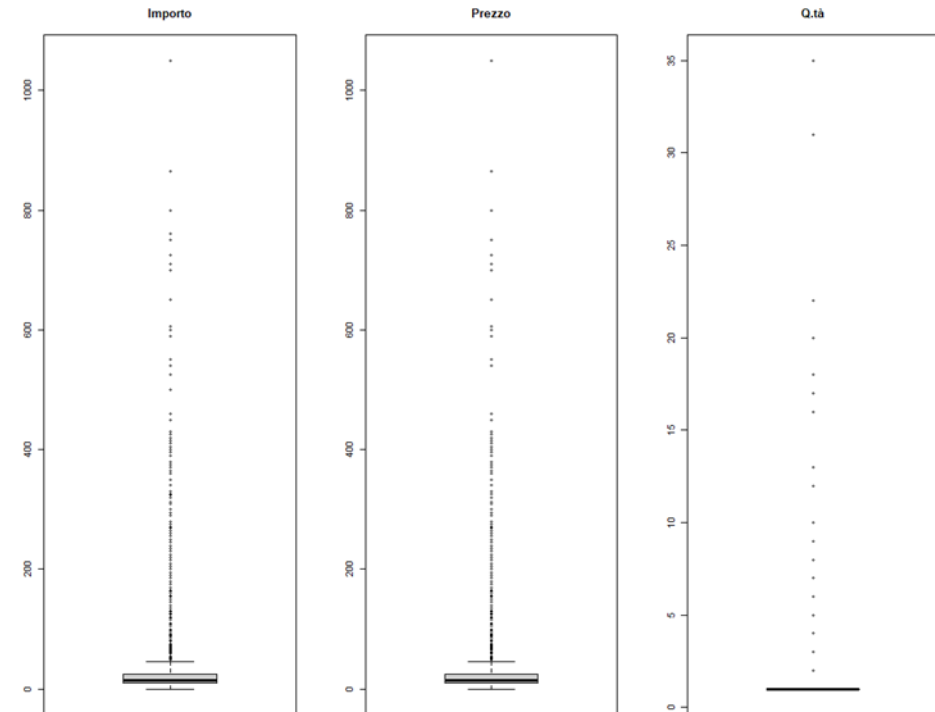
```
par(mfrow=c(1,3))
```

```
boxplot(Fatturato$Importo, main="Importo")
```

```
boxplot(Fatturato$Prezzo, main="Prezzo")
```

```
boxplot(Fatturato$Q.tà, main="Q.tà")
```

I Boxplot evidenziano  
l'asimmetria della  
distribuzione delle variabili.

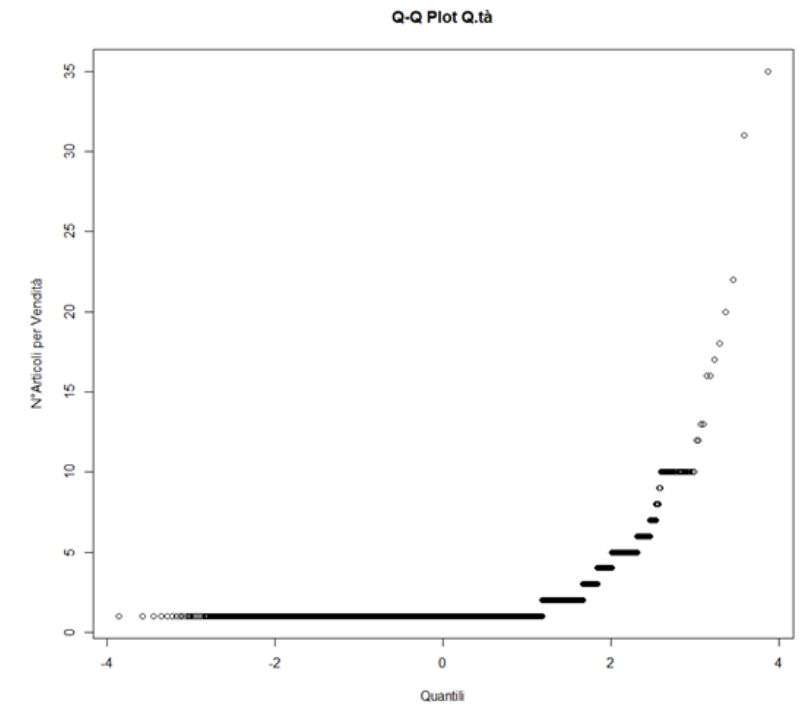
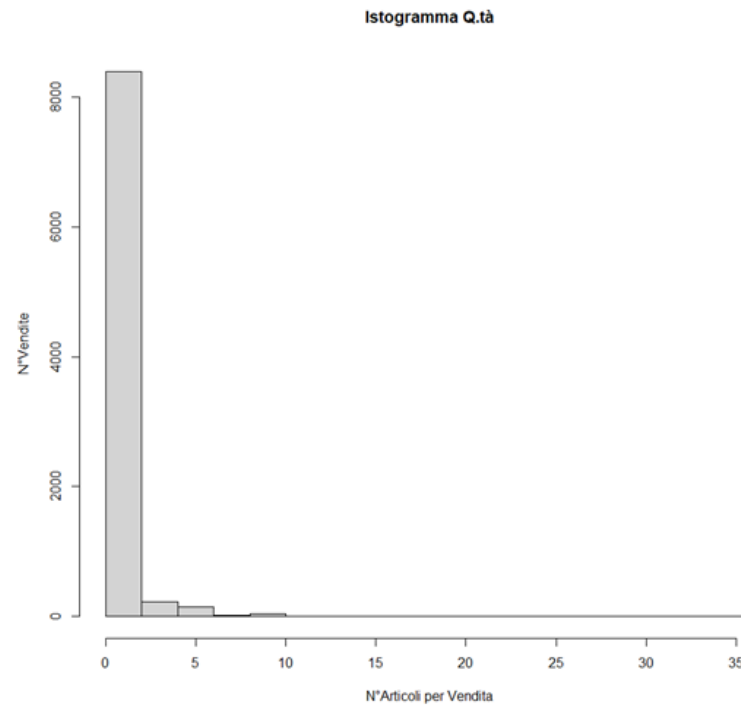


## ANALISI DEL DATASET

### Analisi descrittiva univariata

Studio la distribuzione delle variabili quantitative Q.tà e Fatturato:

```
par(mfrow=c(1,2))  
hist(Fatturato$Q.tà, main="Istogramma Q.tà",  
      xlab="N ° Articoli per Vendita", ylab="N ° Vendite")  
qqnorm(Fatturato$Q.tà, main="Q-Q Plot Q.tà",  
        ylab="N ° Articoli per Vendita", xlab="Quantili")
```



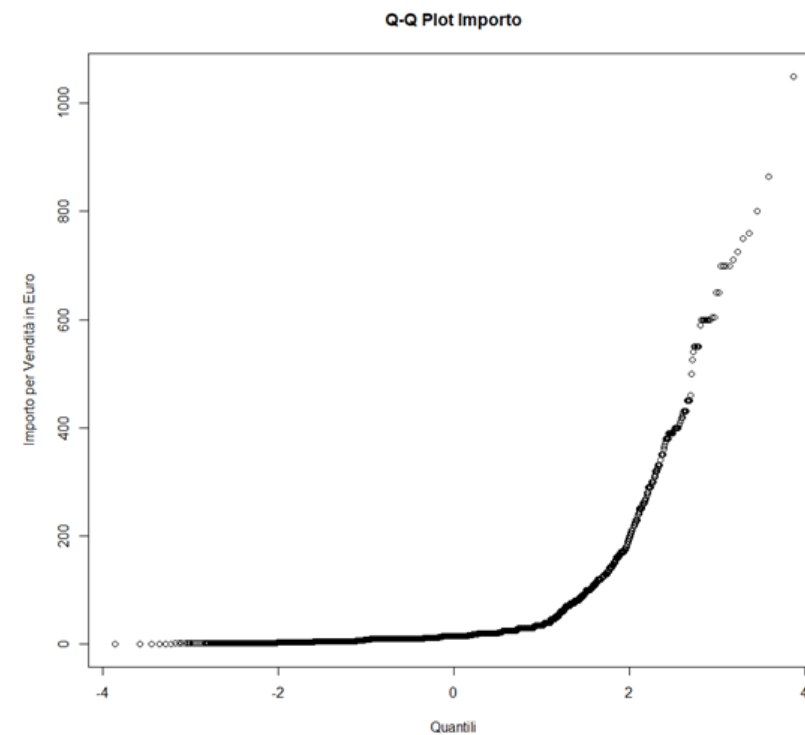
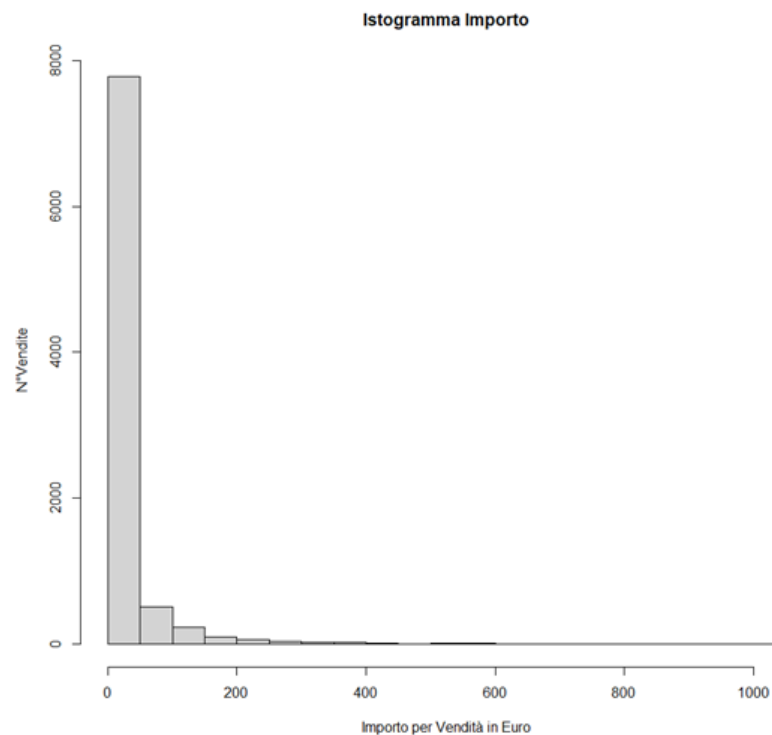


## ANALISI DEL DATASET

### Analisi descrittiva univariata

Entrambe le variabili si discostano ampiamente dalla distribuzione di una normale: dall'istogramma di frequenza si nota una palese asimmetria negativa, cioè le variabili sono maggiormente concentrate sui valori bassi, inoltre entrambe le variabili sono un conteggio, e quindi hanno come supporto gli interi non negativi. Dal QQ-plot noto di nuovo che sia la variabile Q.tà che la variabile Importo non assumono le determinazioni di una normale.

```
par(mfrow=c(1,2))  
hist(Fatturato$Importo, main="Istogramma Importo",  
      ylab="N° Vendite", xlab="Importo per Vendita in Euro")  
qqnorm(Fatturato$Importo, main="Q-Q Plot Importo",  
        ylab="Importo per Vendita in Euro", xlab="Quantili")
```

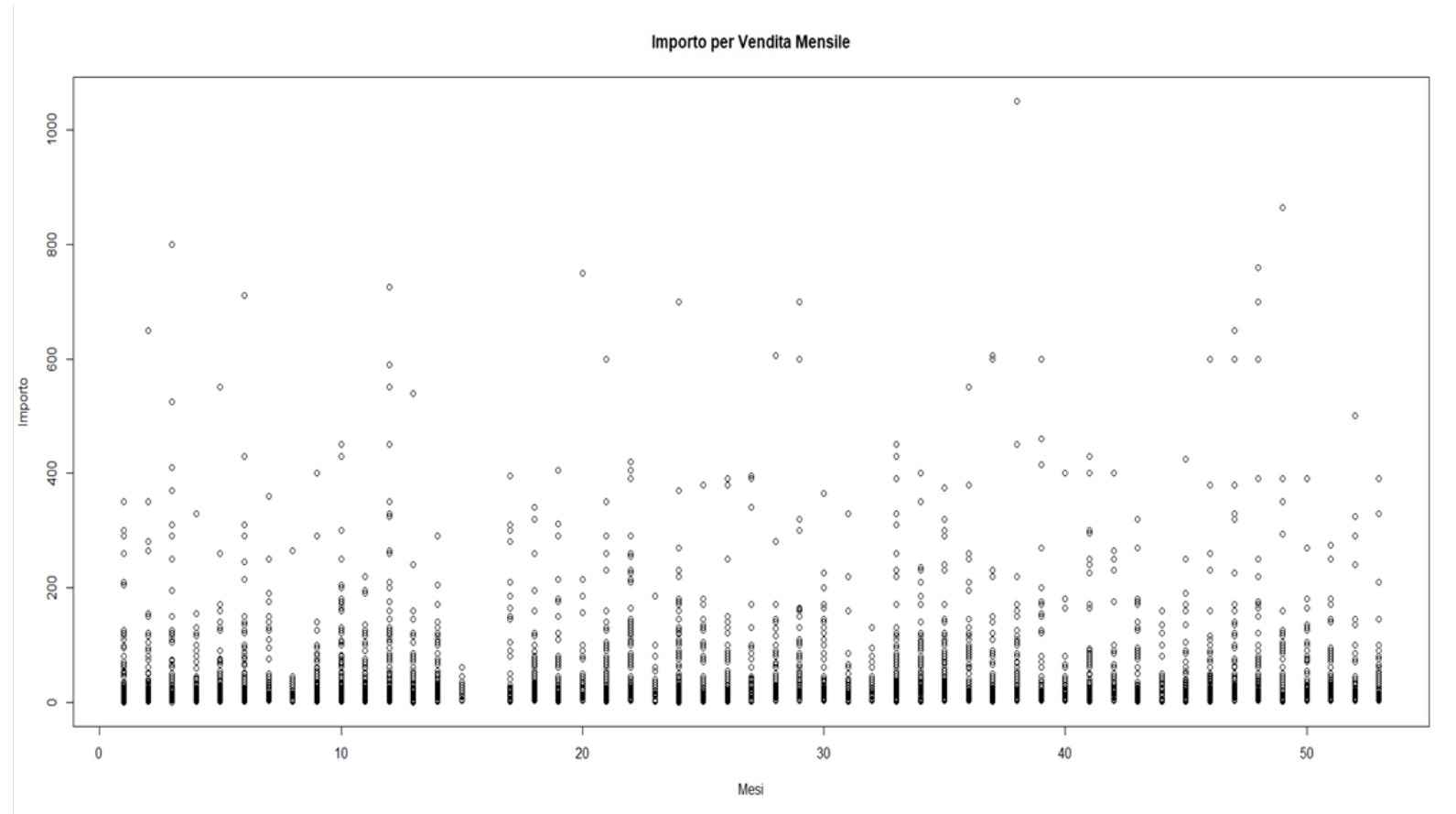


## ANALISI DEL DATASET

### Analisi descrittiva bivariata

Conduco le analisi descrittive delle variabili contenute nel dataset Fatturato:

```
plot(Fatturato$Mese_Anno, Fatturato$Importo,  
      xlab="Mesi", ylab="Importo", main="Importo per Vendita Mensile")
```



## ANALISI DEL DATASET

### Analisi descrittiva bivariata

Approfondisco quindi l'analisi delle vendite andando a guardare la variabile Categoria che suddivide gli articoli del negozio in 5 serie: Accessori, Audio, Libri, Liuteria e Strumenti.

```
> summary(Accessori)
Q.tà      Fatturato
Min.   : 0.0   Min.   : 0
1st Qu.: 85.0  1st Qu.:1169
Median :121.0  Median :1748
Mean   :125.4  Mean   :1624
3rd Qu.:157.0  3rd Qu.:2005
Max.   :290.0  Max.   :3240
```

```
> summary(Audio)
Q.tà      Fatturato
Min.   : 0.0   Min.   : 0.0
1st Qu.:24.0  1st Qu.: 609.0
Median :32.0  Median : 812.0
Mean   :34.4  Mean   : 931.8
3rd Qu.:42.0  3rd Qu.:1306.0
Max.   :95.0  Max.   :2834.0
```

```
> summary(Libri)
Q.tà      Fatturato
Min.   : 0.00  Min.   : 0.0
1st Qu.: 7.00  1st Qu.:118.9
Median :15.00  Median :224.5
Mean   :17.96  Mean   :254.6
3rd Qu.:24.50  3rd Qu.:365.2
Max.   :66.00  Max.   :813.0
```

```
> summary(Liuteria)
Q.tà      Fatturato
Min.   : 0.00  Min.   : 0.0
1st Qu.:13.00  1st Qu.: 216.0
Median :19.00  Median : 400.0
Mean   :19.15  Mean   : 438.1
3rd Qu.:25.00  3rd Qu.: 583.0
Max.   :40.00  Max.   :1038.0
```

```
> summary(Strumenti)
Q.tà      Fatturato
Min.   :0.00  Min.   : 0
1st Qu.: 7.00  1st Qu.:1100
Median :13.00  Median :1776
Mean   :14.58  Mean   :1983
3rd Qu.:18.00  3rd Qu.:2669
Max.   :53.00  Max.   :6565
```

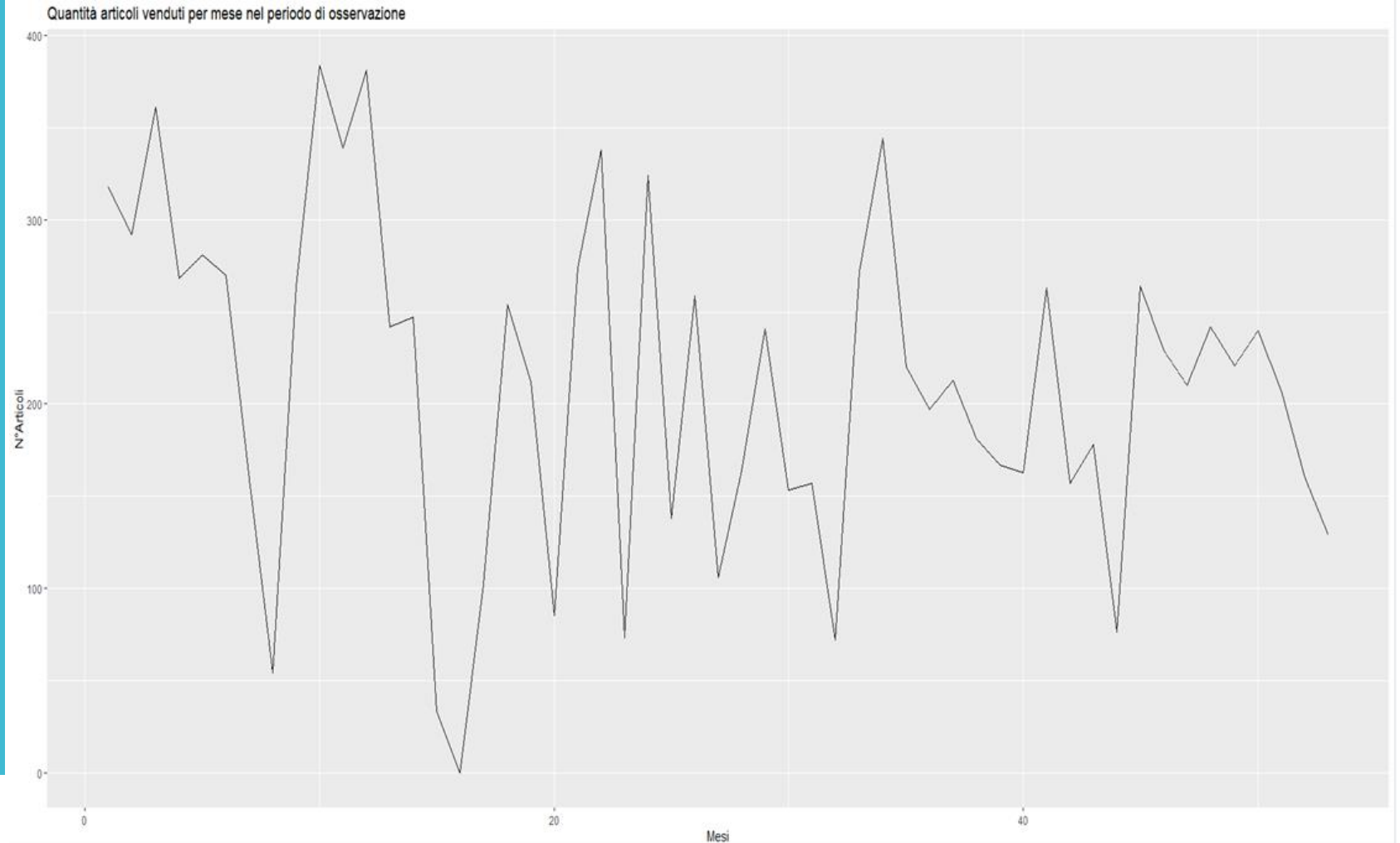
Per ogni serie, per mese, si vede il valore minimo e massimo (Q.ta e Fatturato), la media, la mediana e i quartili. Sono informazioni interessanti, ad esempio osservando la serie Accessori si evidenzia che il valore massimo di Fatturato per mese è stato di 3240€ e il numero massimo di Accessori venduti in un mese è stato di 290 articoli. La mediana ci dice che i clienti del negozio mensilmente acquistano 121 Accessori per un Importo pari a 1748€ .

Analogamente possiamo discutere tutte le altre serie.

Osservando la serie Strumenti si evidenzia che il valore massimo di Fatturato per mese è stato di 6565€ e il numero massimo di strumenti venduti in un mese è stato di 53 articoli.

La mediana ci dice che i clienti del negozio mensilmente acquistano 13 Strumenti per un Importo pari a 1776€.

Le mediane del Fatturato/mese della serie Accessori e della serie Strumenti sono molto vicine. È evidente che gli Strumenti costano circa 9 volte più degli accessori, ma la tendenza mediana del Fatturato per mese è la stessa.



## SERIE STORICHE

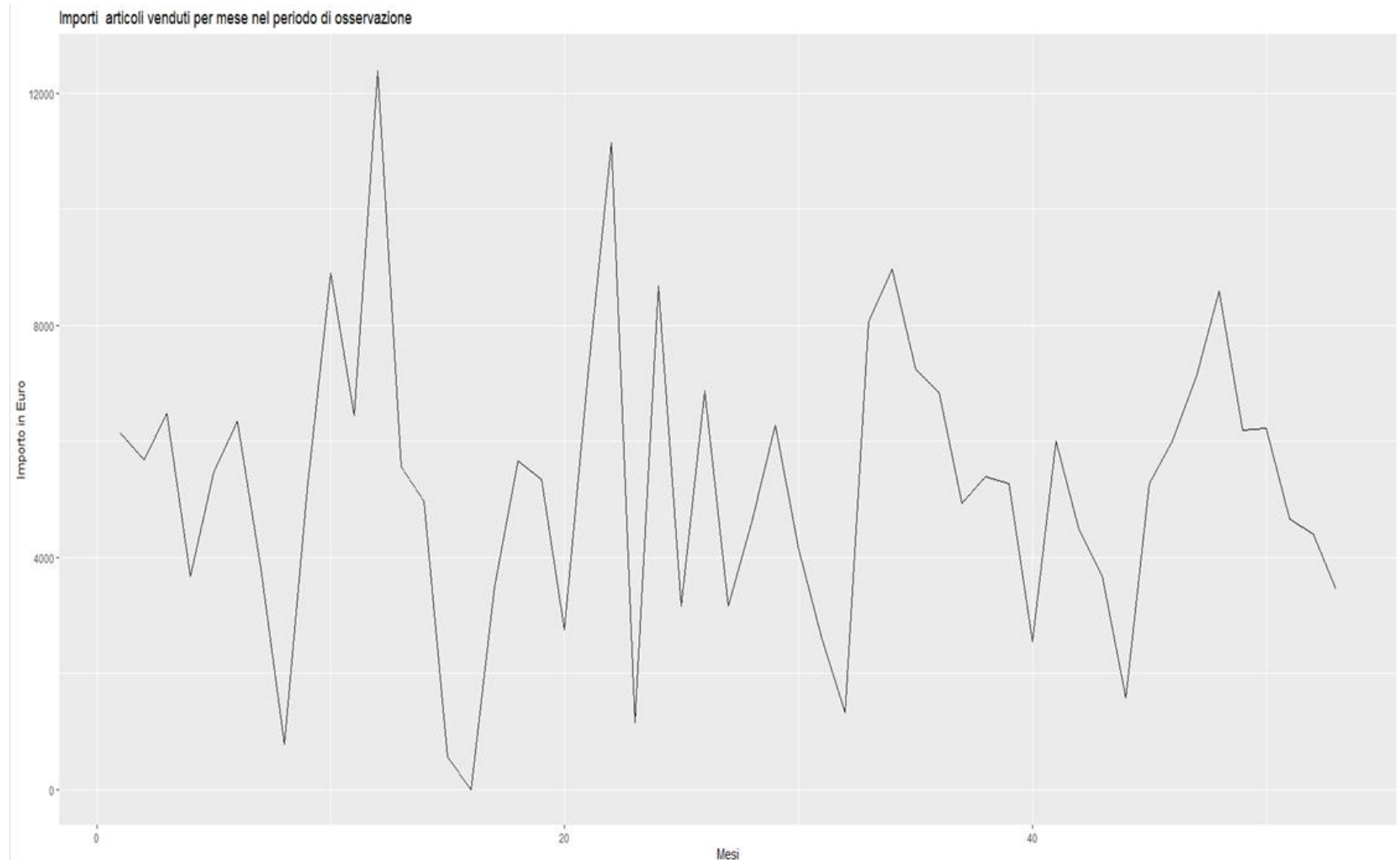
In entrambi i grafici si evidenzia un andamento non lineare; essendo i dati a cadenza mensile, in essi è presente il fenomeno denominato stagionalità. Si può notare che i picchi elevati si registrano sempre nei mesi caldi (novembre e dicembre) e vicini alle feste natalizie.

Si nota inoltre che entrambe le variabili hanno un andamento nel tempo molto simile e pur essendo grandezze di misura diverse i due grafici hanno degli andamenti comuni.

## 2. Serie Importo

Analizzo la variabile Importo

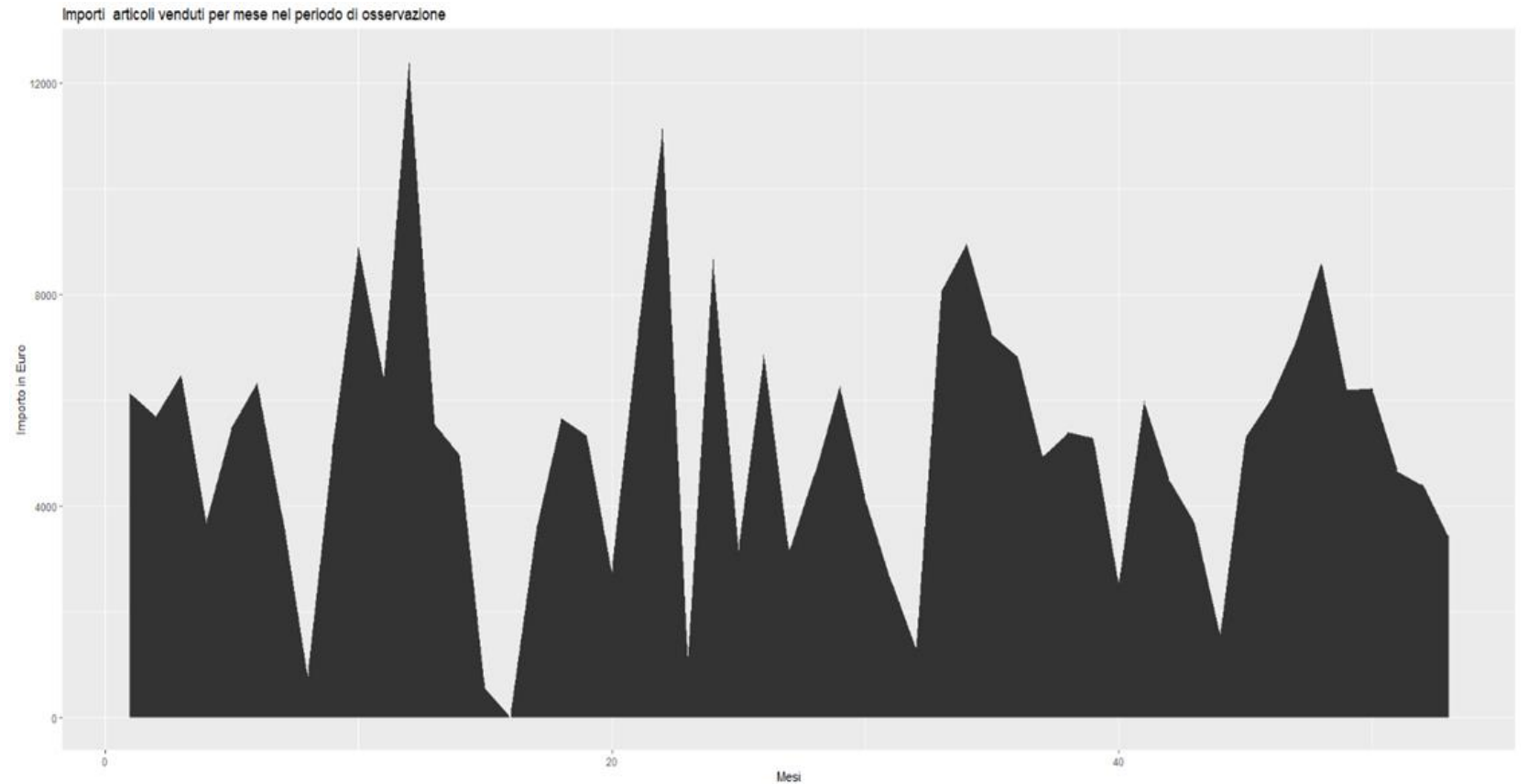
```
ggplot(data = TotSum, aes(x = Mese_Anno, y = Fatturato)) +  
  geom_line() + ggtitle("Importi articoli venduti per mese nel periodo di osservazione") +  
  xlab("Mesi") + ylab("Importo in Euro")
```



## SERIE STORICHE

Una rappresentazione alternativa è quella data dal grafico sottostante che evidenzia l'area e rende più chiara l'eventuale stagionalità

```
ggplot(data = TotSum, aes(x = Mese_Anno, y = Fatturato)) +  
  geom_area() + ggtitle("Importi articoli venduti per mese nel periodo di osservazione") +  
  xlab("Mesi") + ylab("Importo in Euro")
```



## SERIE STORICHE

### Andamenti tipici delle serie storiche

Dai grafici delle serie storiche visti deduciamo che queste possono avere diversi pattern.

1. Pattern stagionale. Questo esiste quando la serie è influenzata da fattori stagionali (es. mensile, semestrale, trimestrale, ecc.). Le serie influenzate dalla stagionalità sono dette anche serie periodiche poiché il ciclo stagionale si ripete in un periodo fisso. Nei dati di tipo annuale la stagionalità non è presente. Dai grafici delle serie storiche vediamo che le variabili Q.tà e Fatturato presentano delle punte principalmente nei mesi 12, 24, 36 e 48. È evidente la stagionalità.
2. Pattern ciclico. Questo tipo di andamento è presente quando la serie presenta aumenti e diminuzioni che non sono di periodo fisso. Questa è la principale differenza fra le fluttuazioni cicliche e quelle stagionali. Inoltre, l'ampiezza delle oscillazioni cicliche è generalmente più grande di quella dovuta alla stagionalità. Il pattern ciclico è determinato dalle espansioni e contrazioni dell'economia dovuti a fenomeni congiunturali. Nei grafici delle serie storiche notiamo alcune oscillazioni cicliche che andremo ad analizzare. (vedi la serie Liuteria)
3. Trend o tendenza di fondo. È caratterizzato da un andamento crescente o decrescente di lungo periodo.

Vedremo di seguito analizzando le diverse serie della variabile Categoria i diversi tipi di pattern.

## RELAZIONI TRA SERIE

Vado alla ricerca di una qualche relazione tra le variabili presenti nel dataset studiandole congiuntamente.

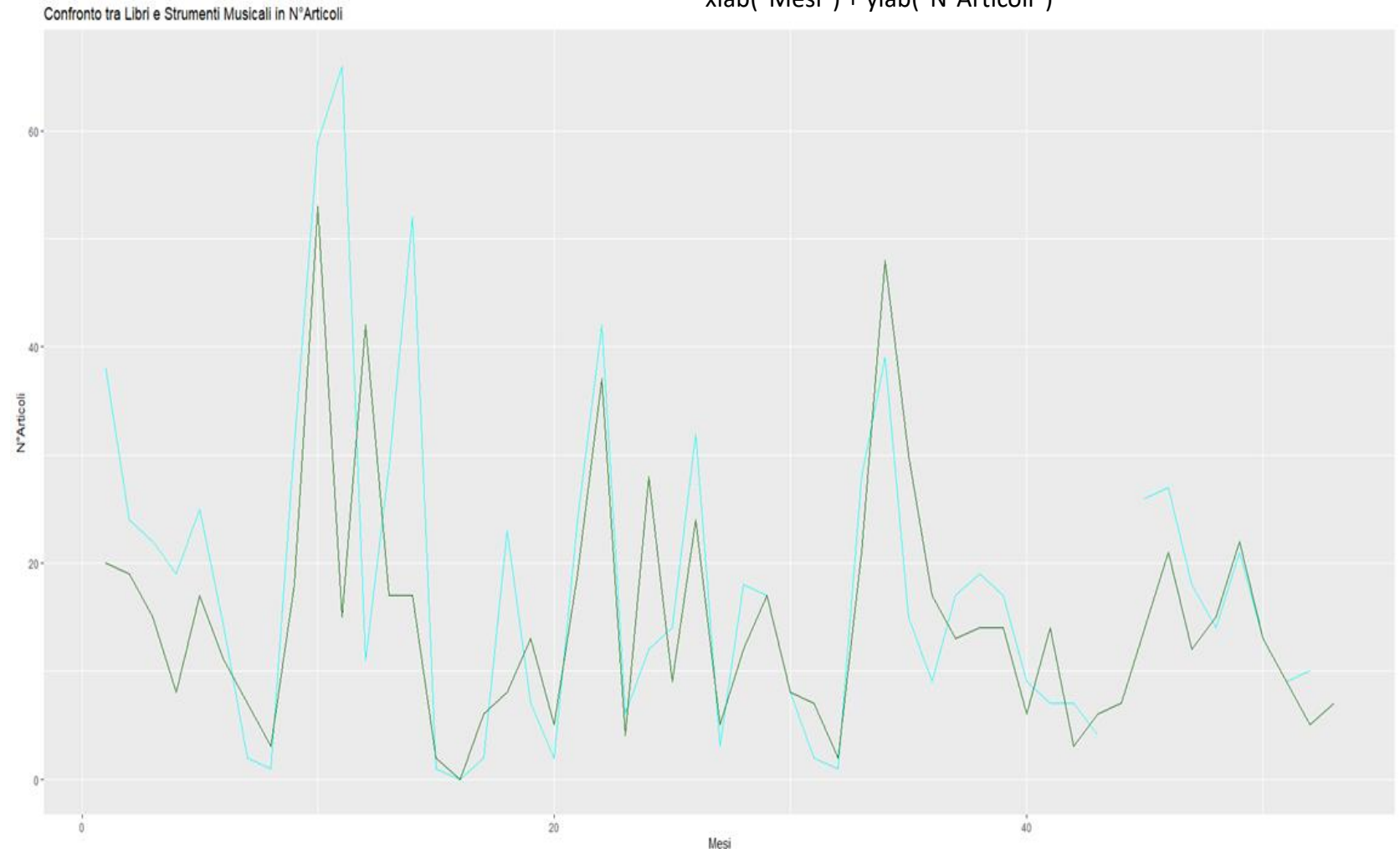
Per far ciò riporto sullo stesso grafico gli andamenti temporali di 2 variabili.

Nel 1° caso analizzo l'andamento Q.tà / mese della serie Libri confrontandola con la serie Strumenti.

Si nota che le due serie hanno un andamento molto simile, quasi sovrapponibile. In questo caso, è possibile ipotizzare che a fronte di ogni strumento acquistato, viene acquistato ugualmente un libro.

### 1. Relazione tra serie Libri e Strumenti

```
ggplot(CattPezz, aes(x=Mese_Anno)) +  
  geom_line(aes(y = LIBRI), color = "cyan") +  
  geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +  
  ggtitle("Confronto tra Libri e Strumenti Musicali in N°Articoli") +  
  xlab("Mesi") + ylab("N°Articoli")
```

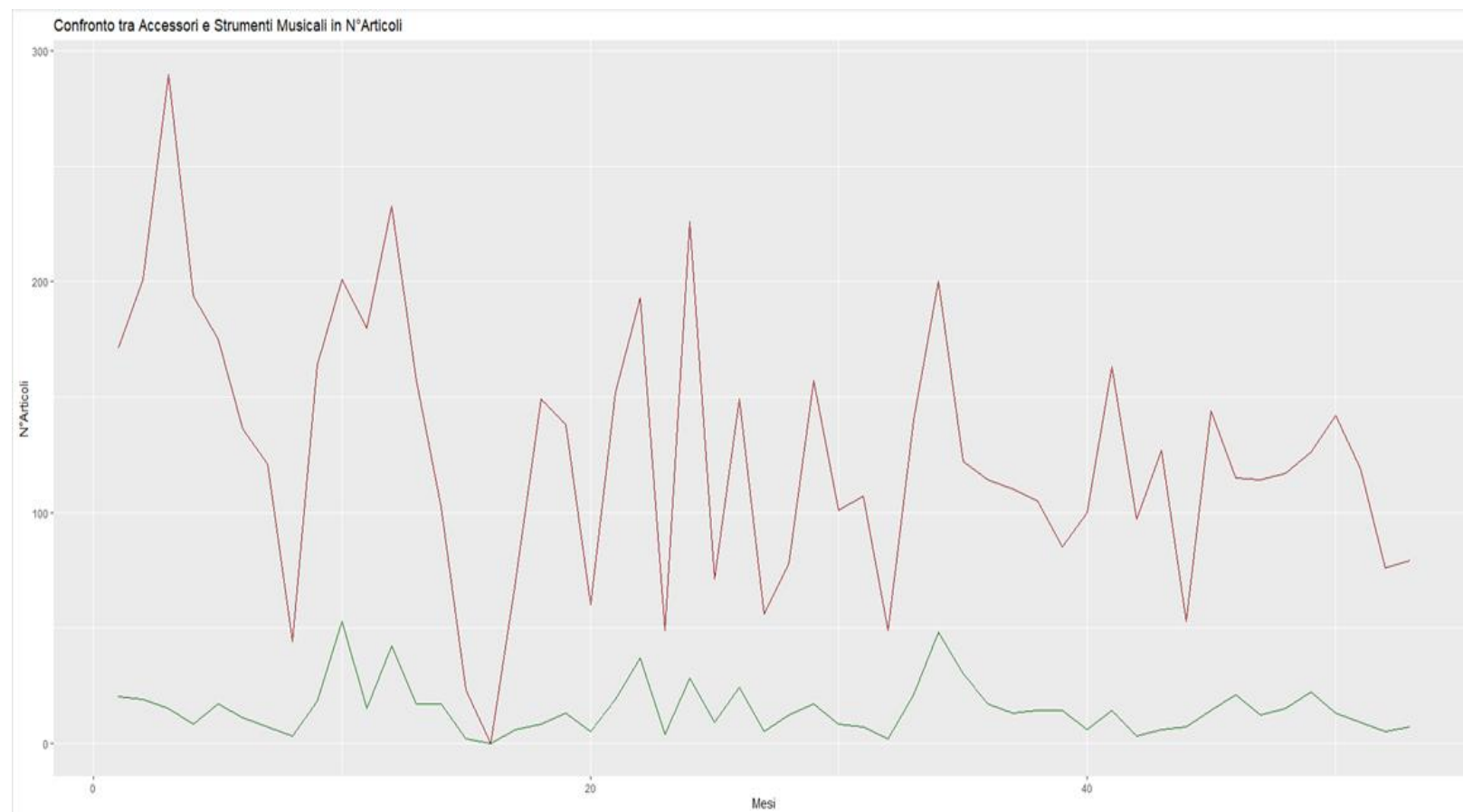




## RELAZIONI TRA SERIE

### 2. Relazione tra serie Accessori e Strumenti Musicali

```
ggplot(CattPezz, aes(x=Mese_Anno)) +  
  geom_line(aes(y = ACCESSORI), color = "darkred") +  
  geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +  
  ggtitle("Confronto tra Accessori e Strumenti Musicali in N°Articoli") +  
  xlab("Mesi") + ylab("N°Articoli")
```

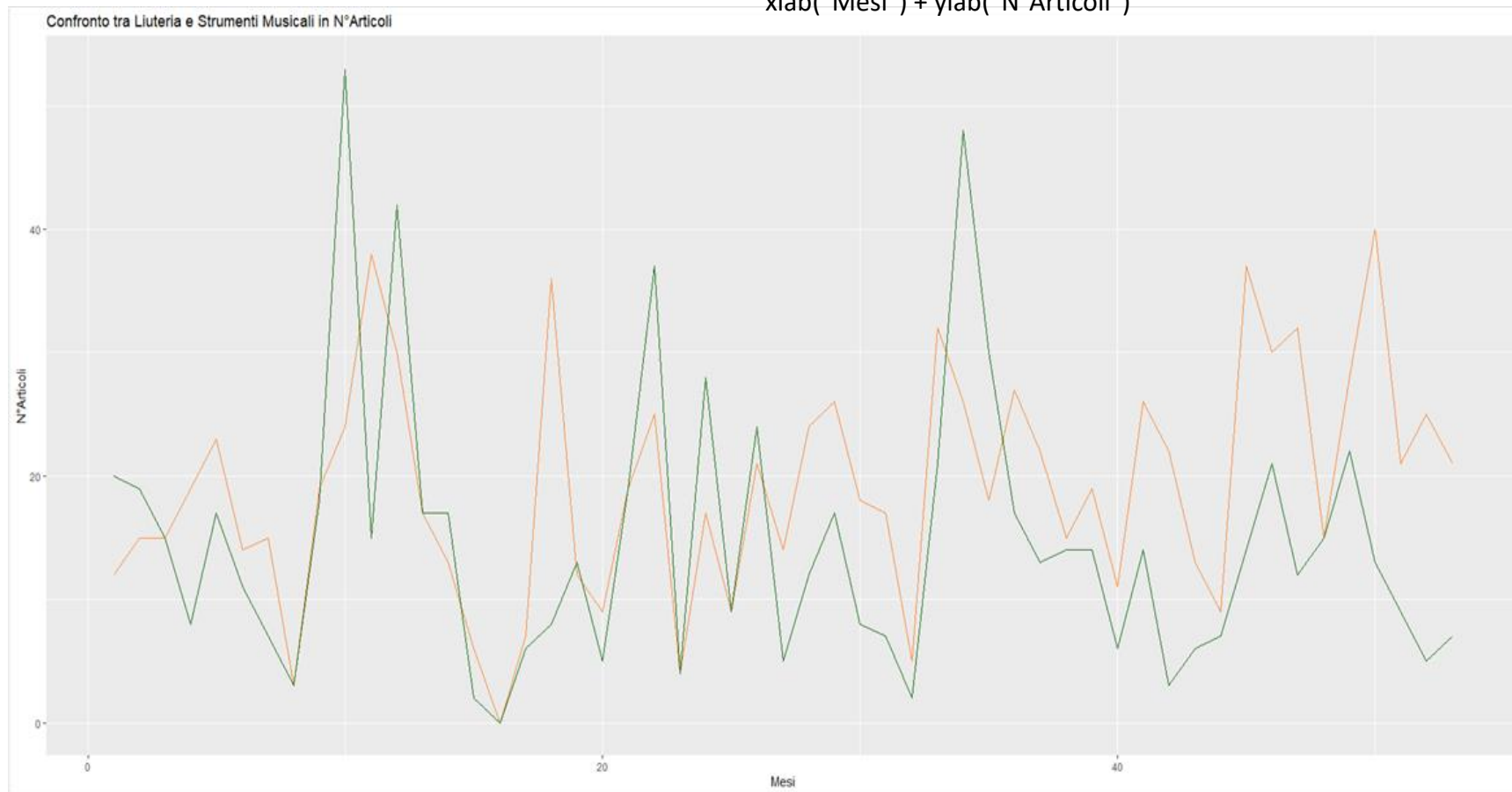


## RELAZIONI TRA SERIE

Questo caso è piuttosto particolare. La serie Liuteria ha un pattern stagionale, ma anche un pattern ciclico. Questo tipo di andamento è presente quando la serie presenta aumenti e diminuzioni che non sono di periodo fisso. Inoltre, l'ampiezza delle oscillazioni cicliche è generalmente più grande di quella dovuta alla stagionalità. Nelle serie economiche il pattern ciclico è determinato dalle espansioni e contrazioni dell'economia dovuti a fenomeni congiunturali.

### 3. Relazione tra serie Liuteria e Strumenti

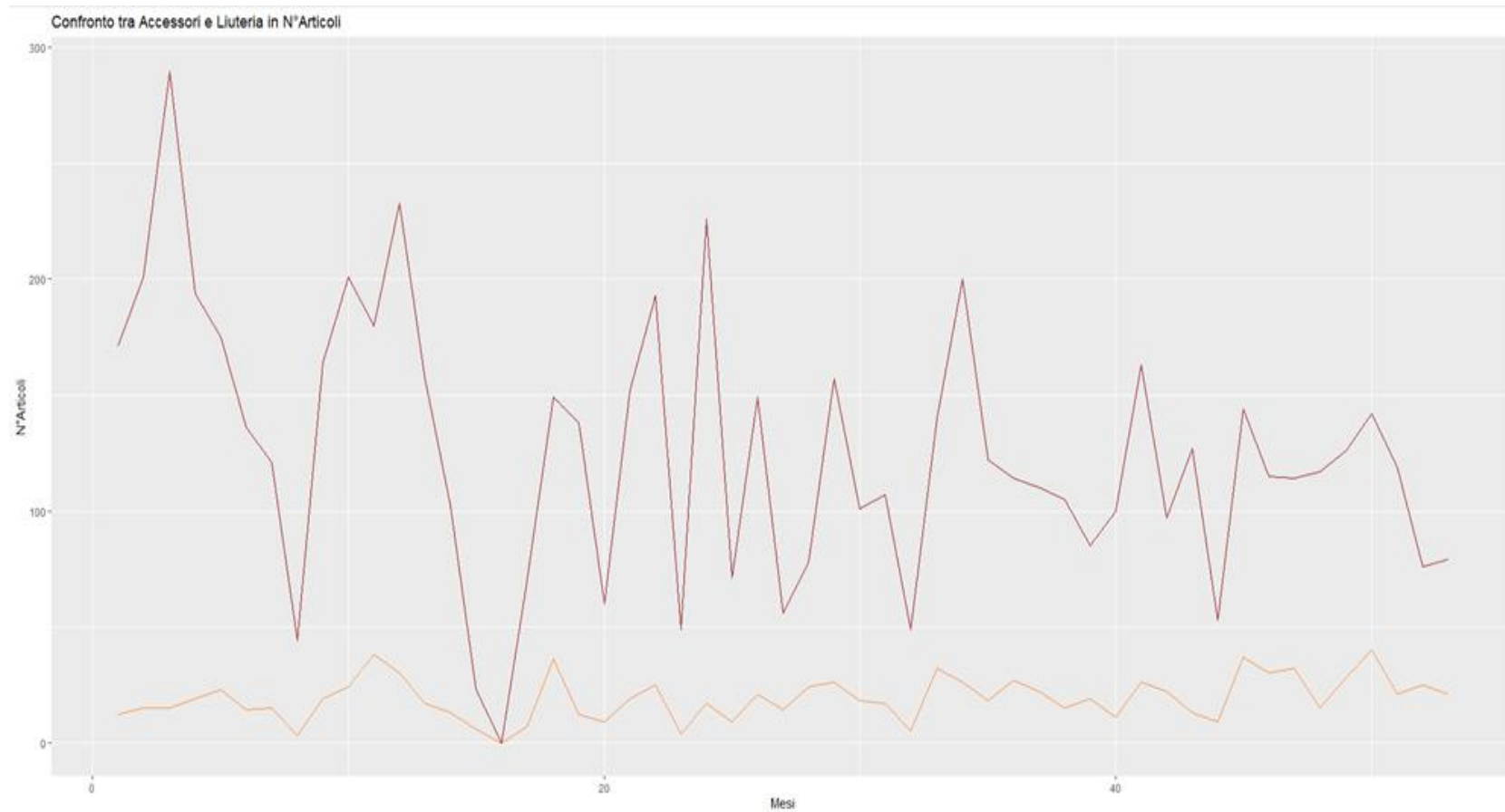
```
ggplot(CattPezz, aes(x=Mese_Anno)) +  
  geom_line(aes(y = LIUTERIA), color = "chocolate1") +  
  geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +  
  ggtitle("Confronto tra Liuteria e Strumenti Musicali in N°Articoli") +  
  xlab("Mesi") + ylab("N°Articoli")
```



## RELAZIONI TRA SERIE

### 4. Relazione tra serie Accessori e Liuteria

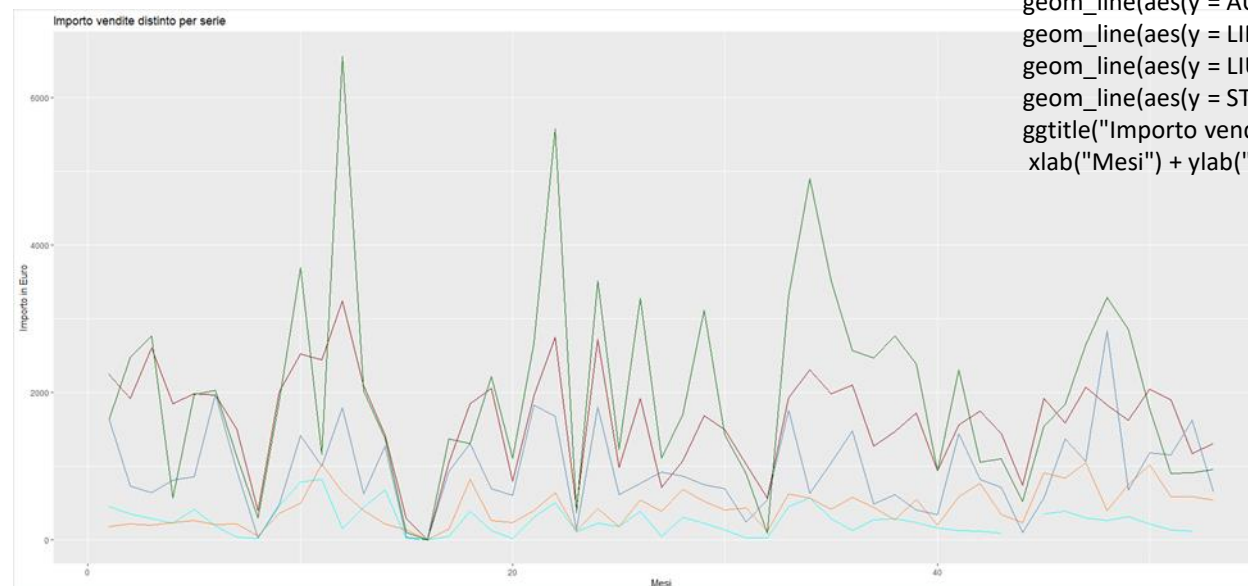
```
ggplot(CattPezz, aes(x=Mese_Anno)) +  
  geom_line(aes(y = ACCESSORI), color = "darkred") +  
  geom_line(aes(y = LIUTERIA), color = "chocolate1") +  
  ggtitle("Confronto tra Accessori e Liuteria in N°Articoli") +  
  xlab("Mesi") + ylab("N°Articoli")
```



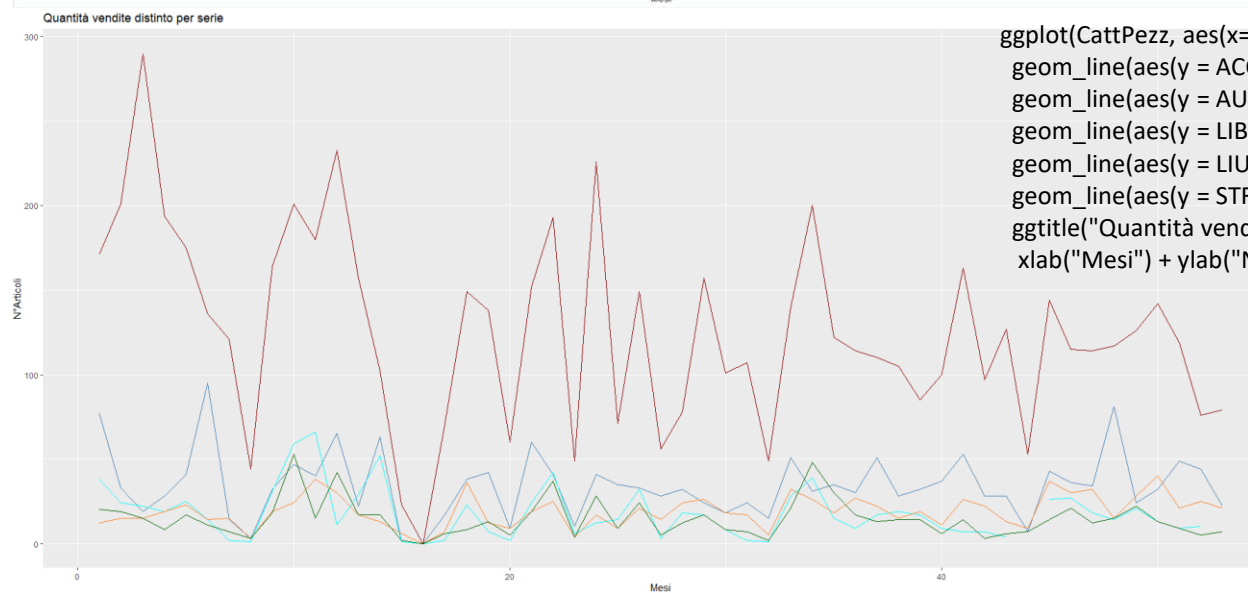
## RELAZIONI TRA SERIE

Guardando sia il numero Q.tà che gli Importi per ogni sottogruppo della variabile Categoria si nota che il sottogruppo accessori è quello in assoluto più venduto, ma la caratteristica della stagionalità si rileva sempre in tutti i sottogruppi.

## 5. Relazione tra le 5 serie



```
ggplot(FattCatt, aes(x=Mese_Anno)) +  
  geom_line(aes(y = ACCESSORI), color = "darkred") +  
  geom_line(aes(y = AUDIO), color = "steelblue") +  
  geom_line(aes(y = LIBRI), color = "cyan") +  
  geom_line(aes(y = LIUTERIA), color = "chocolate1") +  
  geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +  
  ggtitle("Importo vendite distinto per serie") +  
  xlab("Mesi") + ylab("Importo in Euro")
```

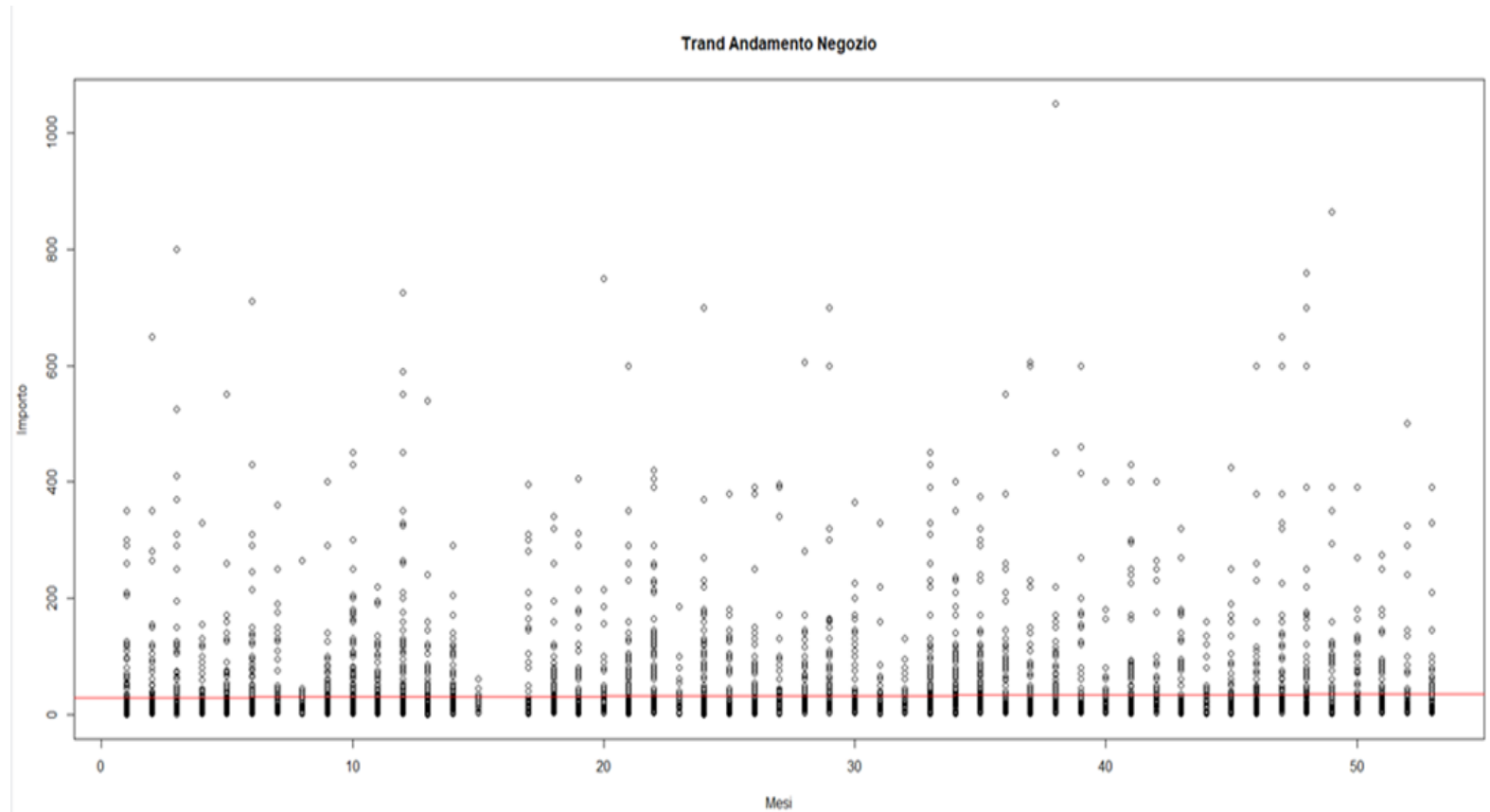


```
ggplot(CattPezz, aes(x=Mese_Anno)) +  
  geom_line(aes(y = ACCESSORI), color = "darkred") +  
  geom_line(aes(y = AUDIO), color = "steelblue") +  
  geom_line(aes(y = LIBRI), color = "cyan") +  
  geom_line(aes(y = LIUTERIA), color = "chocolate1") +  
  geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +  
  ggtitle("Quantità vendite distinto per serie") +  
  xlab("Mesi") + ylab("N°Articoli")
```

## RELAZIONI TRA SERIE

### 4. Trend Fatturato Negozio

```
data(Fatturato)
plot(data = Fatturato, Importo ~ Mese_Anno,
     ylab = "Importo", xlab="Mesi", main="Trand Andamento Negozio")
abline(lm(data = Fatturato, Importo ~ Mese_Anno), col = "red")
```



## CONCLUSIONE

La mia analisi ha evidenziato che serie storiche delle variabili Q.tà e Fatturato presentano una combinazione di diversi pattern delle vendite.

Abbiamo visto che i grafici delle serie storiche delle variabili Q.tà e Fatturato presentano delle punte nei mesi 12, 24, 36 e 48 (dicembre - Festività Natalizie), delle punte leggermente più basse nei mesi 5, 17, 29, 41, 53 (maggio – Comunioni e Cresime) e delle cadute nei mesi 8, 20, 32, 44 (chiusura parziale del mese di agosto). E' evidente la stagionalità. Anche le 5 serie Audio, Accessori, Libri, Liuteria e Strumenti sono influenzate dal pattern stagionale.

La serie Liuteria è un esempio di pattern ciclico. Si vede che l'ampiezza delle oscillazioni cicliche è generalmente più grande di quella dovuta alla stagionalità.

L'ultimo grafico evidenzia un andamento nel Fatturato del negozio costante; visto questo dato costante nei primi 4 anni e 5 mesi di vita del negozio non ho ritenuto necessario sviluppare un modello predittivo.

Grazie dell'attenzione.