

# Analisi dei dati relativi ai risultati dell' Eurovision song contest

Campo Lorenzo

2022-07-01

## Cos'è l'Eurovision song contest?

L'eurovision song copntest è un concorso musicale a livello internazionale che si svolge ogni anno nel mese di Maggio.

Fondato dai Big 5 nel 1956, è il concorso musicale più longevo al mondo

I Big 5 o grandi 5 sono i 5 stati fondatori degli eurovision e sono Regno Unito, Spagna, Francia, Germania e Italia. Lo status di Big 5 garantisce la qualifica istantanea alle finali. Per questo motivo le nostre analisi si concentreranno principalmente sui risultati delle finali poiché in caso contrario, stati che hanno partecipato alle semifinali si ritroverebbero con più punti di quanti gli spetterebbero

Lo stato ospitante è lo stato che ha ottenuto la vittoria l'anno precedente

## Come avviene la votazione?

La valutazione delle canzoni viene divisa tra giuria e televoto. Ogni stato assegna voti alle 10 canzoni più piaciute in base alla classifica stilata dal televoto e dalla giuria della propria nazione in modo separato. Ogni nazione NON può votare per se stessa I punti assegnati a ogni canzone sono assegnati in questo modo

- primo posto:12 punti
- secondo posto 10 punti
- terzo posto in poi: da 8 a 1 punto

Il massimo punteggio complessivo è di 24 per ogni stato votante meno il proprio

Si noti che questo sistema di votazione è relativamente recente (introdotto nel 2016). Eurovision ha cambiato molto spesso i suoi sistemi di votazione e, come vedremo in seguito, ciò influirà sull'analisi dei dati.

Ora che abbiamo capito di cosa stiamo parlando possiamo cominciare l'analisi

## Librerie e importazione del dataset

Cominciamo con l'importare le librerie che ci serviranno per la nostra analisi

```
#http://www.esstats.com/ fonte dei dati  
#https://data.world/rhubarbarosa/eurovisionvotingstats Download dataset
```

```
library("dplyr")  
library("readr")  
library(tidyr)  
library(ggplot2)  
library(purrr)  
library(broom)  
library(stringr)  
library(modelr)
```

```
options(scipen = 999) #ci serve ad eliminare la notazione scientifica
```

Importiamo ora 2 dei dataset che useremo durante la nostra analisi - events\_data: contiene i dati generali dell'evento e sarà la base della maggior parte dei nostri data frame. - event\_televote: contiene dati dettagliati delle votazioni dal 2016 al 2018 con i dettagli della votazione di giuria e televoto e relativi piazzamenti in classifica -event\_country\_votes: mostra stati votanti e stati riceventi per ogni votazione della giuria

Il nostro dataset contiene dati a partire dalla prima edizione del 1956 fino al 2019 ma la prima edizione, poiché in quell'occasione il sistema di votazione non dichiarava il punteggio finale, verrà ignorato in quanto non contiene dati utili ai nostri scopi e potrebbe influire sui risultati di certi grafici

Inoltre il 2019 è presente solo in un dataset ma, poiché non contiene dati che possono gravemente influenzare gli esiti dei nostri esami, ho ritenuto opportuno tenerlo in quanto idoneo nel contesto preso in considerazione

```
events_data <- read_delim("E:/dataset eurovision/EventParticipantsAndPlace.csv", delim = ";", escape_double = FALSE)  
events_data=events_data%>%  
  mutate(Country=replace(Country,Country=="Itali","Italy"))%>%  
  mutate(Country=replace(Country,Country=="Sqitzerland","Switzerland"))%>%  
  arrange(event.year)%>%  
  filter(event.year>1956)  
  
event_televote<-read_delim("E:/dataset eurovision/VotingDetails.csv", delim = ";", escape_double = FALSE)  
  
event_country_vote<-read_delim("E:/dataset eurovision/CountryVotes.csv", delim = ";", escape_double = FALSE)  
  
state_votes<-event_country_vote%>%  
  filter(event.type=="final" | event.type=="grand-final")%>%  
  select(event.year,awarding.country,recipient.country,pts)
```

## Tabelle di supporto

Creiamo ora 2 tabelle che ci serviranno più avanti. Queste tabelle sono: - `edition_max_votes`: fornisce il massimo punteggio attribuito da uno stato a un altro in base all'edizione - `support`: contiene il numero di partecipanti in ogni edizione.

```
event.year=c(1957:2018)
i=0
max_points=rep(0,length(event.year))
for (var in event.year) {
  i=i+1
  x=0
  if(var<=1961)
  {
    x=10
  }
  else
  if(var==1962)
  {
    x=3
  }
  else
  if(var>=1963 & var<=1966)
  {
    x=5
  }
  else
  if(var>=1967 & var<=1973)
  {
    x=10
  }
  else
  if(var==1974)
  {
    x=10
  }
  else
  if(var>=1975 & var<=2015)
  {
    x=12
  }
  else
  if(var>=2016 & var<=2022)
  {
    x=24
  }
  max_points[i]=x
}

edition_max_votes<-data.frame(event.year,max_points)
support=events_data%>%
  filter(event.year>1956)%>%
  group_by(event.year)%>%
  summarize(partecipanti=n_distinct(Country))
```

## statistiche interessanti

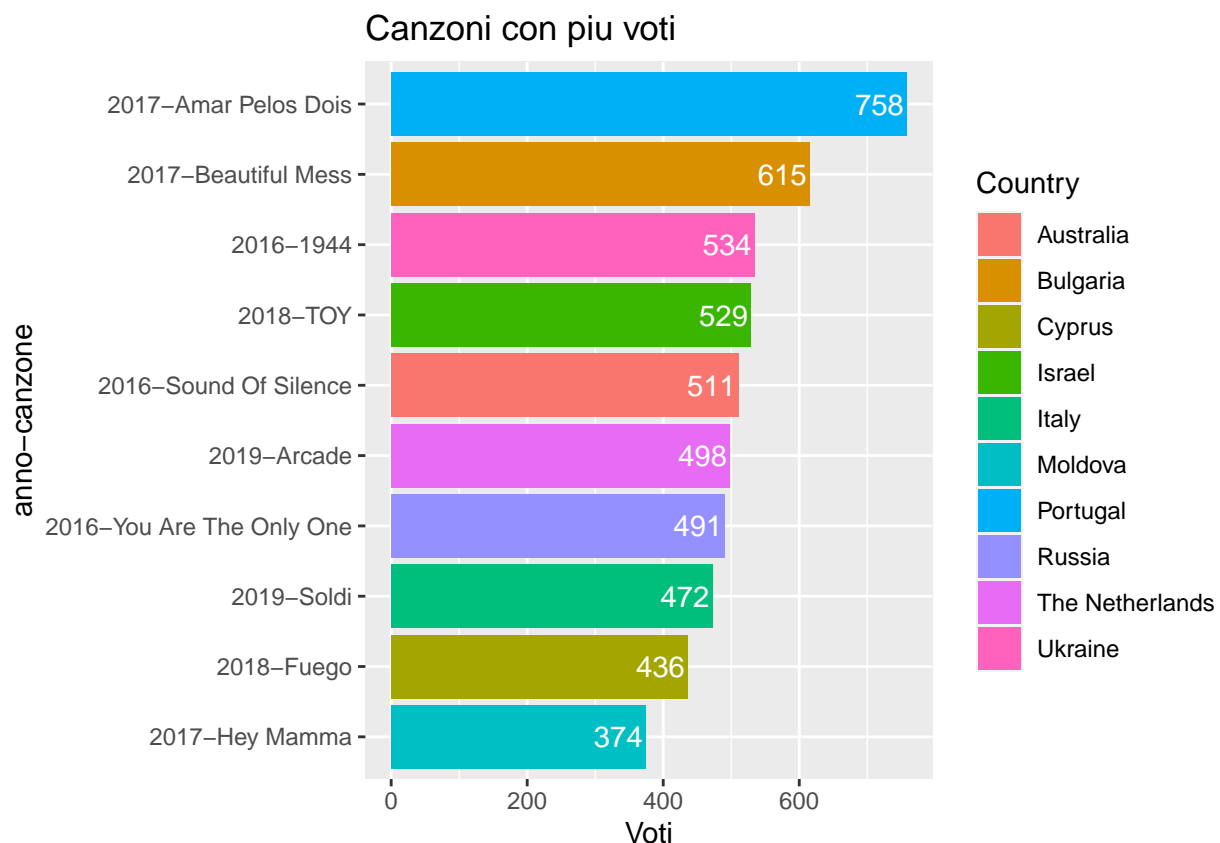
Possiamo ora cominciare a rispondere ad alcune delle nostre domande

### Quali canzoni hanno ricevuto piu voti?

Quali sono state le canzoni piu votate nella storia del festival? Ci limitiamo ad analizzare i dati a partire dal 2016 poiché, come spiegato in precedenza, coincide con l'entrata in vigore del nuovo sistema di punteggio e i dati delle edizioni precedenti non sono sufficienti a elaborare una conversione accurata.

```
most_voted_songs<-events_data%>%
  filter(event.type=="final" | event.type=="grand-final",event.year>=2016)%>%
  mutate(Entry=paste(event.year,Song))%>%
  select(Entry,Country,pts)%>%arrange(-pts)

ggplot(head(most_voted_songs,10),mapping=aes(y=reorder(str_replace(string=Entry,pattern="[:graph:]]"),
  geom_col()+
  labs(title="Canzoni con piu voti",
    x="Voti", y= "anno-canzone")+
  geom_text(aes(label = pts),nudge_x = -40, color="white")
```



Congratulazioni a Amar Pelos Dois, canzone del portogallo del 2017, per essere la canzone piu apprezzata nella storia recente degli eurovision

“Perche non valutare la percentuale di punti ricevuti sul totale possibile? In questo modo si potrebbero valutare anche le edizioni precedenti”

Semplice, perché non conosciamo quanti hanno votato e quanti di questi voti siano legittimi (sono sorti spesso casi di voti concordati, 5 solo quest'anno), ma solo il totale.

## Quale nazione Ha totalizzato piu vittorie?

Un conto è ricevere un buon punteggio, ma un altro è ottenere l'agoniato premio. Vediamo ora chi è stato lo stato piu abile

```
most_victory<-events_data%>%filter(event.type=="final" | event.type=="grand-final",Place==1)%>%group_by  
most_victory
```

```
## # A tibble: 27 x 2  
##   Country      vittorie  
##   <chr>         <int>  
## 1 Ireland         7  
## 2 Sweden          6  
## 3 France          5  
## 4 Luxembourg      5  
## 5 The Netherlands 5  
## 6 United Kingdom  5  
## 7 Israel          4  
## 8 Denmark         3  
## 9 Norway          3  
## 10 Austria        2  
## # ... with 17 more rows
```

```
ggplot(data=most_victory ,mapping=aes(y= reorder(Country,vittorie),x=vittorie))+  
  geom_col(fill="#ffd700")+  
  labs(title="Nazioni con più vittorie",  
        x="vittorie", y= "Stato")
```



Congratulazioni all'irlanda per aver totalizzato ben 7 vittorie seguita da Svezia e Regno unito

## Quale nazione ha totalizzato più ultime posizioni?

Passiamo ora dalla parte opposta della classifica

Chi ha totalizzato più ultime posizioni?

```
Finalists<-events_data%>%  
  filter(event.type=="final" | event.type=="grand-final")%>%  
  select(event.year,Place)%>%group_by(event.year)%>%summarise(finalisti=max(Place))
```

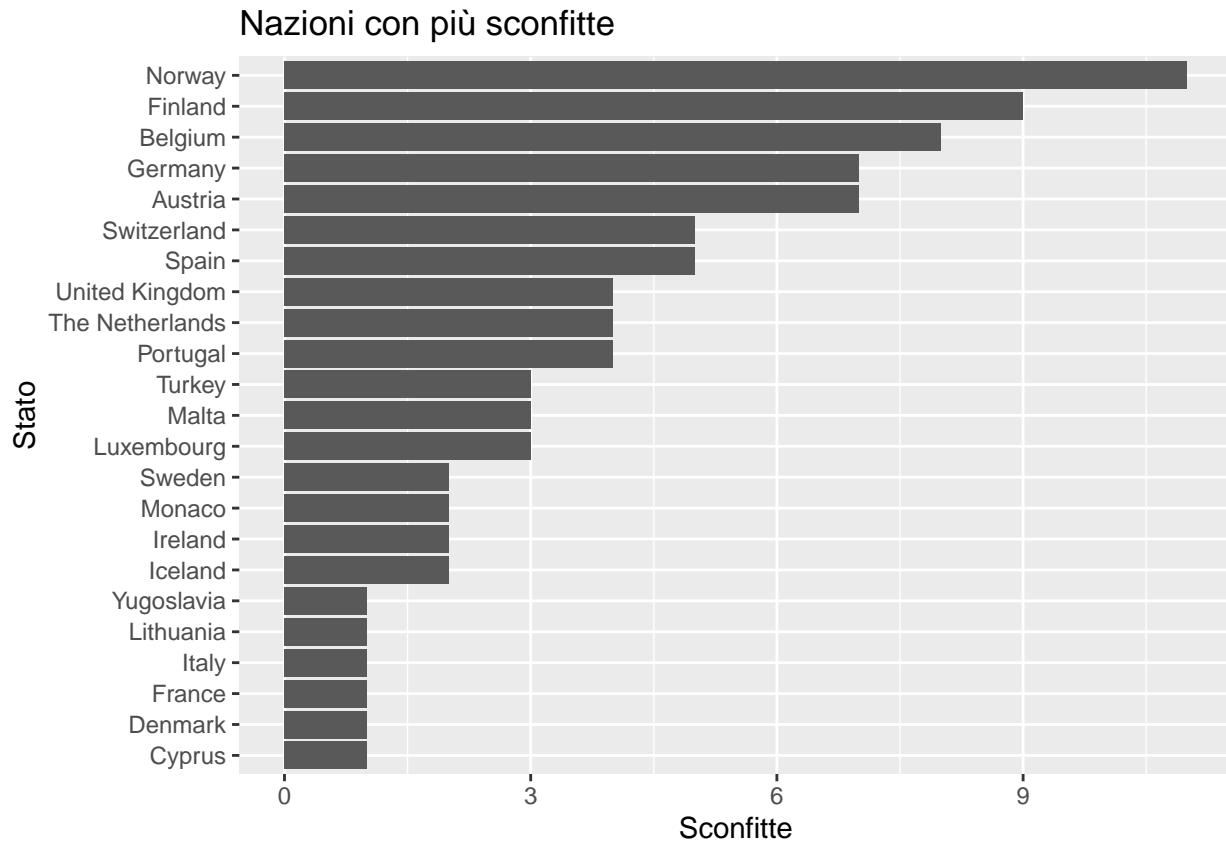
```
most_last<- inner_join(events_data,Finalists)%>%filter(event.type=="final" | event.type=="grand-final"),
```

```
## Joining, by = "event.year"
```

```
sum(most_last[2])
```

```
## [1] 87
```

```
ggplot(data=most_last ,mapping=aes(y= reorder(Country,ultimi),x=ultimi))+  
  geom_col()+  
  labs(title="Nazioni con più sconfitte",  
        x="Sconfitte", y= "Stato")
```



Purtoppo questo triste encomio viene guadagnato dalla Norvegia che totalizza un totale di 11 sconfitte totali Seguono Belgio con 10 e Germania a parimerito con la Finlandia con 9 ultime posizioni



## Quale la nazione ha collezionato piu seconde posizioni?

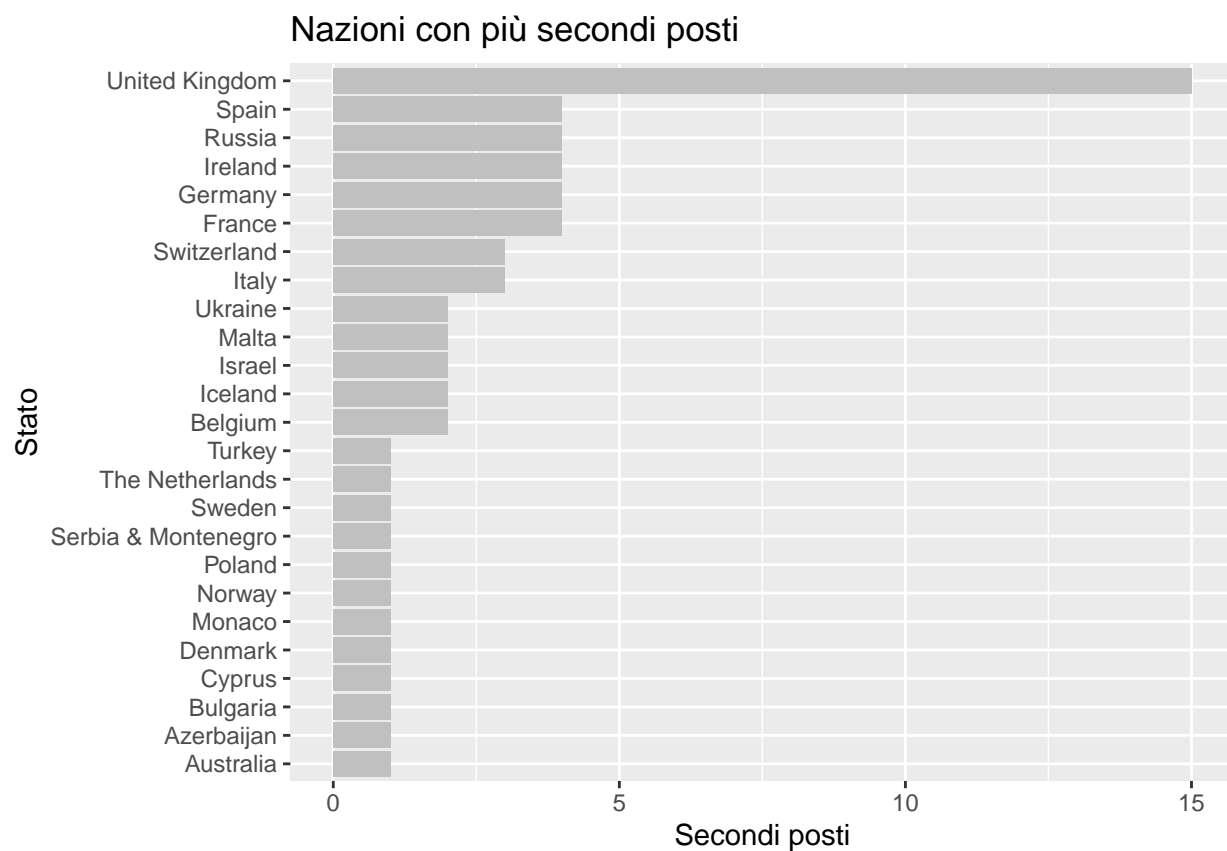
Non sempre il primo classificato ottiene tutta la gloria. Spesso i “runner up” o secondi classificati acquisiscono più successo dei loro colleghi al primo posto e spesso appaiono nuovamente in edizioni successive del festival stesso.

Vediamo chi è quasi arrivato alla vetta

```
most_second<-events_data%>%filter(event.type=="final" | event.type=="grand-final",Place==2)%>%arrange(Country,runner_up)
most_second
```

```
## # A tibble: 25 x 2
##   Country      runner_up
##   <chr>         <int>
## 1 United Kingdom      15
## 2 France              4
## 3 Germany             4
## 4 Ireland             4
## 5 Russia              4
## 6 Spain              4
## 7 Italy               3
## 8 Switzerland        3
## 9 Belgium            2
## 10 Iceland           2
## # ... with 15 more rows
```

```
ggplot(data=most_second ,mapping=aes(y= reorder(Country,runner_up),x=runner_up))+
  geom_col(fill="#c0c0c0")+
  labs(title="Nazioni con più secondi posti",
        x="Secondi posti", y= "Stato")
```



Con un enorme distacco il Regno Unito si guadagna il premio di eterno secondo con un totale di 15 secondi posti Seguono Germania, Russia, Irlanda, Germania e Francia con 4 secondi posti l'uno

Quali sono gli stati che hanno guadagnato più punti, quanti di questi sono stati attribuiti dal televoto e quanti dalla giuria?

Ora facciamo un piccolo confronto vediamo ora rappresentato quanto hanno contribuito televoto e giuria nelle votazioni dei vari stati.

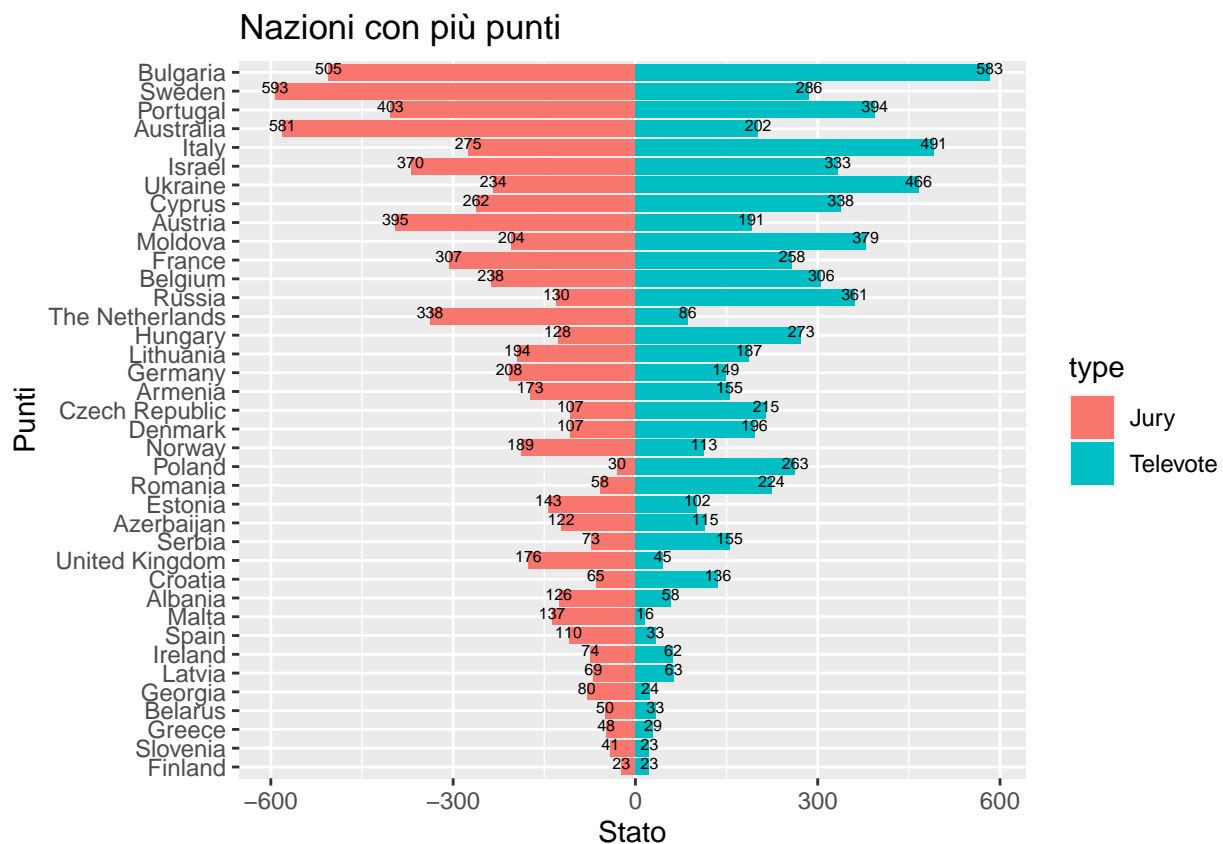
```
piramid_j<-event_televote%>%filter(event.type=="final" | event.type=="grand-final")
names(piramid_j)[names(piramid_j) == 'Jury.points'] <- 'pts'

piramid_t<-event_televote%>%filter(event.type=="final" | event.type=="grand-final")
names(piramid_t)[names(piramid_t) == 'Televoting.points'] <- 'pts'

piramid_jt<-full_join(piramid_j,piramid_t)%>%
  group_by(Receiving.country,type)%>%
  summarise(pts=sum(pts,na.rm = TRUE),Total.pts=sum(Total.pts,na.rm = TRUE))%>%
  mutate(pts=ifelse(type=="Jury", pts*(-1), pts*1))%>%arrange(-Total.pts)

## Joining, by = c("event.year", "Receiving.country", "pts", "Total.pts", "type")
## `summarise()` has grouped output by 'Receiving.country'. You can override using
## the `.groups` argument.

ggplot(piramid_jt,aes(x = reorder(Receiving.country>Total.pts),y = pts,fill=type)) +
  geom_col() +
  coord_flip()+
  labs(title="Nazioni con più punti",
       x="Punti", y= "Stato")+
  geom_text(size=2,aes(label = abs(pts)),nudge_x = 0.2, color="black")
```



Vediamo che esistono alcuni casi poco bilanciati come la Polonia e Regno Unito

La polonia deve il suo successo al televoto mentre, al contrario, il regno unito lo deve alla giuria

### Qual'è la posizione media ottenuta da ogni stato?

Vediamo infine la media complessiva dei piazzamenti dei vari stati

```
media_pos<-events_data%>%filter(event.type=="final" | event.type=="grand-final")%>%group_by(Country)%>%  
knitr::kable(media_pos,  
caption = "Media posizione in classifica")
```

Table 1: Media posizione in classifica

Country	media.Place
Serbia & Montenegro	4.500000
Bulgaria	6.250000
Monaco	7.380952
Italy	7.568182
Russia	7.619048
Sweden	8.379310
Azerbaijan	8.727273
Ireland	8.866667
Australia	9.000000
Ukraine	9.133333
United Kingdom	9.193548
Luxembourg	9.194444
Armenia	10.000000
Denmark	10.000000
France	10.065574
The Netherlands	10.180000
Switzerland	10.312500
Israel	10.361111
Yugoslavia	10.666667
Estonia	10.937500
Germany	11.112903
Serbia	11.333333
Greece	11.421053
Belgium	11.540000
Latvia	11.600000
Moldova	11.900000
Norway	11.909091
Malta	11.960000
Austria	12.276596
Croatia	12.333333
Georgia	12.428571
Spain	12.440678
Turkey	12.636364
Romania	13.055556
Bosnia & Herzegovina	13.166667
Cyprus	13.400000
Albania	13.666667
Iceland	13.960000
Czech Republic	14.000000
Portugal	14.047619
North Macedonia	14.444444
Finland	14.777778
Hungary	15.071429
Poland	15.428571

Country	media.Place
Slovenia	15.733333
Montenegro	16.000000
Belarus	17.166667
Lithuania	17.307692
Morocco	18.000000
Slovakia	19.333333
San Marino	21.500000

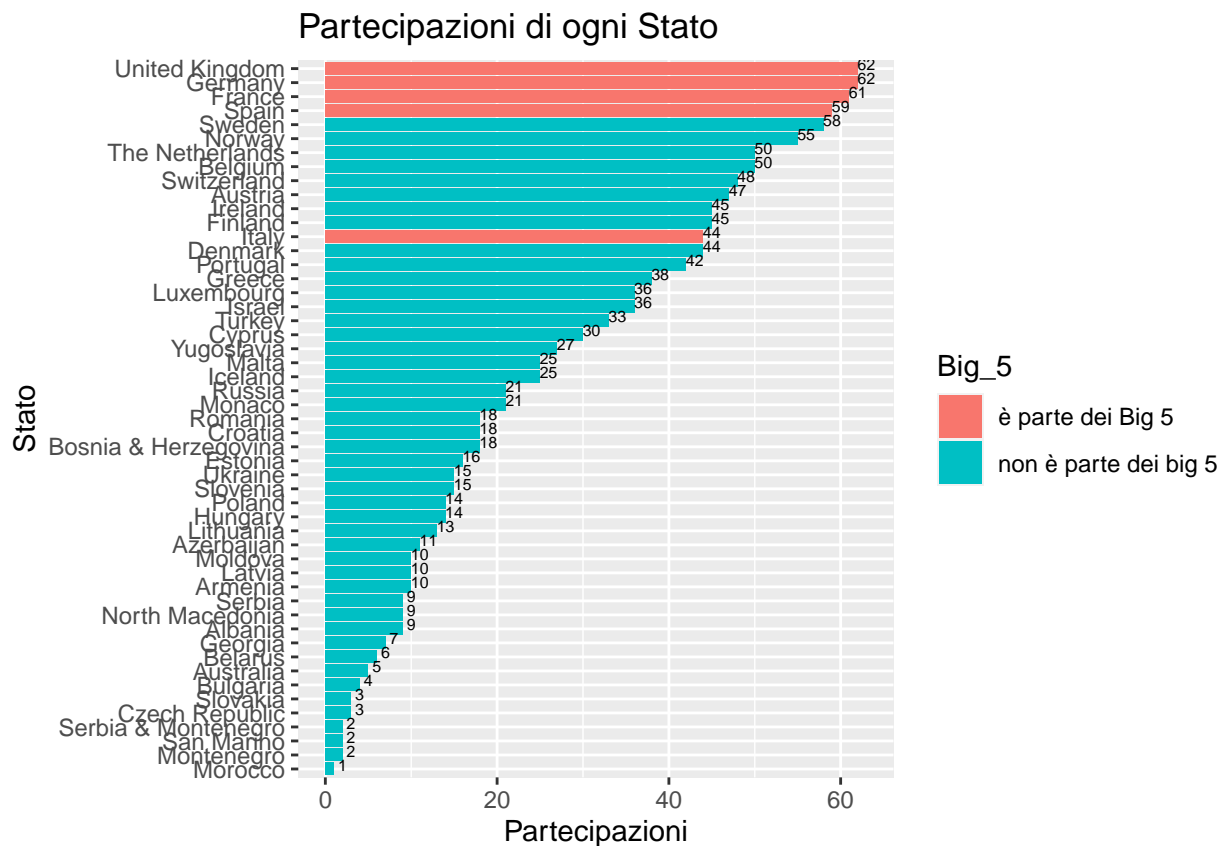
Possiamo notare che al primo posto abbiamo Serbia & Montenegro. Sebbene non siano uno stato a se, il regolamento non vieta a 2 stati di unirsi in un'unica esibizione.

## Quante partecipazioni alla finale dell'eurovision ha guadagnato ogni stato?

Vediamo ora quante volte ogni stato ha partecipato alla finale

```
most_participation<-events_data%>%filter(event.type=="final" | event.type=="grand-final")%>%
  group_by(Country)%>%summarise(partecipazioni=n())%>%arrange(-partecipazioni)%>%mutate(Big_5=ifelse(Co

ggplot(data=most_participation ,mapping=aes(y= reorder(Country,partecipazioni),x=partecipazioni,fill=Bi
  geom_col()+
  labs(title="Partecipazioni di ogni Stato",
        x="Partecipazioni", y= "Stato")+
  geom_text(size=2,aes(label = partecipazioni),nudge_y = 0.3, nudge_x = 1, color="black")
```



Senza troppe sorprese vediamo 4 dei Big 5 in cima alla classifica (Regno unito, Germania, Francia, Spagna) poichè sono automaticamente inclusi alle finali.

L'italia però fa eccezione

## Perche l'italia non ha partecipato in quantità equivalente?

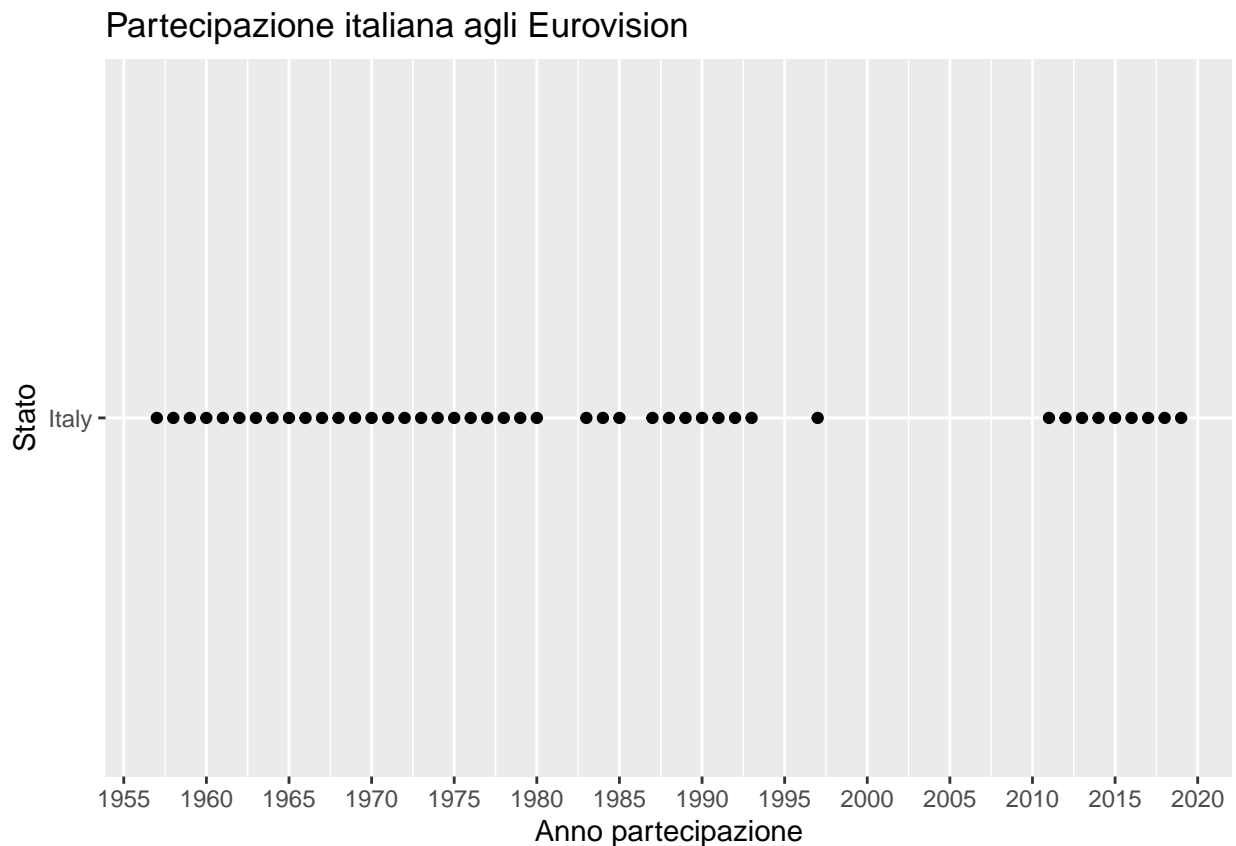
Troviamo il periodo d'attività dell'Italia

```
italy_participation<-events_data%>%filter(event.type=="final" | event.type=="grand-final",Country=="Italy")
  select(event.year,Country)
italy_participation
```

```
## # A tibble: 44 x 2
##   event.year Country
```

```
##          <dbl> <chr>
## 1      1957 Italy
## 2      1958 Italy
## 3      1959 Italy
## 4      1960 Italy
## 5      1961 Italy
## 6      1962 Italy
## 7      1963 Italy
## 8      1964 Italy
## 9      1965 Italy
## 10     1966 Italy
## # ... with 34 more rows
```

```
ggplot(data=italy_participation ,mapping=aes(x=event.year,y=Country))+
  geom_point()+
  scale_x_continuous(breaks = scales::extended_breaks(n=12))+
  labs(title="Partecipazione italiana agli Eurovision",
        x="Anno partecipazione", y= "Stato")
```



Vediamo chiaramente un periodo relativamente esteso di tempo in cui l'italia si è astenuta dal partecipare.

Il motivo è che a quel tempo l'Eurovision non era abbastanza conosciuto e la RAI, emittente che paga i diritti televisivi all'EBU(European Broadcasting Union, prima conosciuta come eurovision, da cui il nome del concorso), non riteneva questo investimento fruttuoso.

L'Italia è tornata nel 2011 dopo continue richieste da parte degli artisti. Da allora gli ascolti dell'eurovision in Italia sono ai massimi storici, specie dopo la vittoria del 2021.



## I Big 5 meritano il loro posto in finale?

Vediamo se il loro risultato medio merita il posto permanente in finale

```
Big_5.pos=media_pos%>%filter(Country %in% c("Italy","Germany","United Kingdom","Spain","France"))  
  
knitr::kable(Big_5.pos,  
  caption = "Media posizione in classifica dei Big 5")
```

Table 2: Media posizione in classifica dei Big 5

Country	media.Place
Italy	7.568182
United Kingdom	9.193548
France	10.065574
Germany	11.112903
Spain	12.440678

Possiamo constatare che mediamente la loro posizione è superiore alla metà dei finalisti (25 minimi in base al regolamento)

Considero l'accesso diretto alla finale come guadagnato.

Quali stati hanno preferenze nel dare i loro 12 punti?

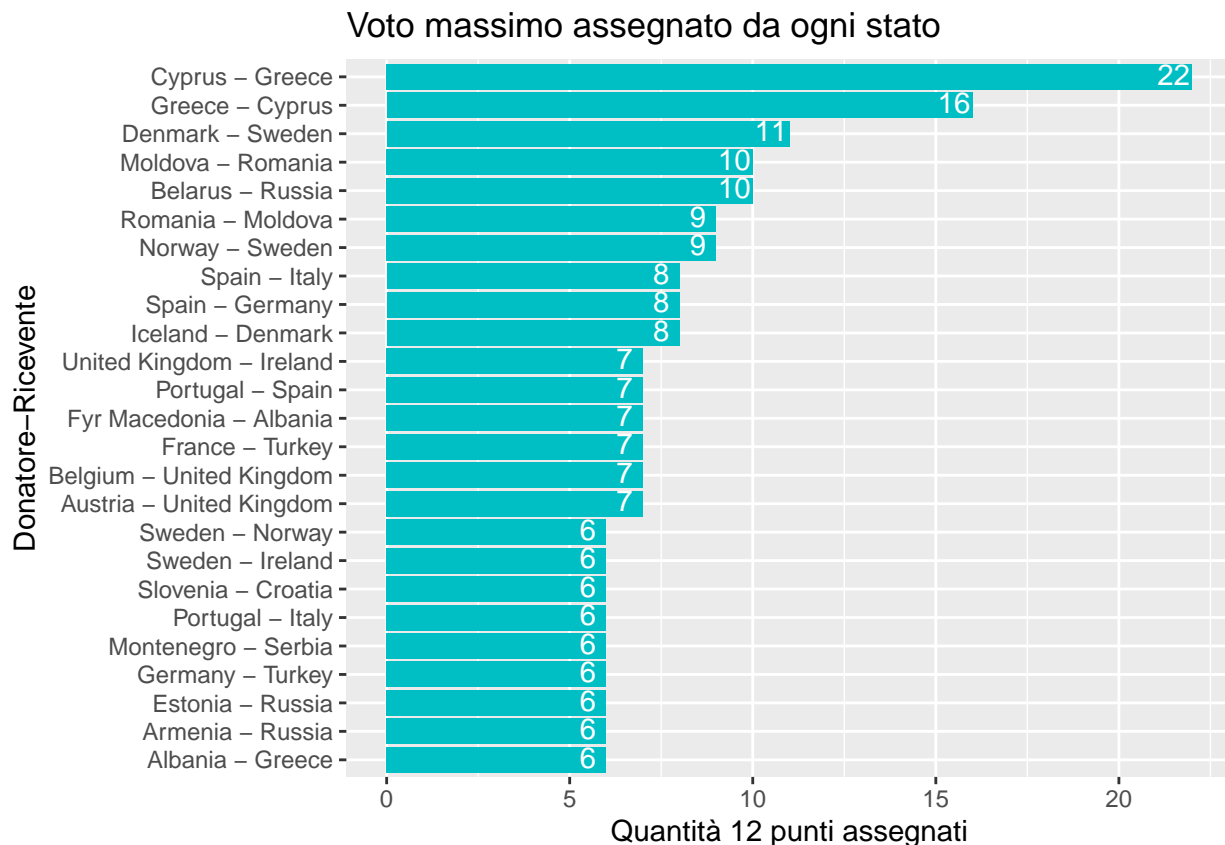
```
most_voted_country<-event_country_vote%>%filter(event.type=="final" | event.type=="grand-final",pts==12)
  group_by(awarding.country,recipient.country)%>%summarise(count=n())%>%arrange(-count)%>%
  mutate(votante_votato=paste(awarding.country,"-",recipient.country))
```

```
## `summarise()` has grouped output by 'awarding.country'. You can override using
## the `.groups` argument.
```

```
most_voted_country
```

```
## # A tibble: 631 x 4
## # Groups:   awarding.country [52]
##   awarding.country recipient.country count votante_votato
##   <chr>           <chr>          <int> <chr>
## 1 Cyprus          Greece           22 Cyprus - Greece
## 2 Greece          Cyprus           16 Greece - Cyprus
## 3 Denmark         Sweden           11 Denmark - Sweden
## 4 Belarus         Russia           10 Belarus - Russia
## 5 Moldova         Romania           10 Moldova - Romania
## 6 Norway          Sweden            9 Norway - Sweden
## 7 Romania         Moldova            9 Romania - Moldova
## 8 Iceland         Denmark            8 Iceland - Denmark
## 9 Spain           Germany            8 Spain - Germany
## 10 Spain          Italy              8 Spain - Italy
## # ... with 621 more rows
```

```
ggplot(data=head(most_voted_country,25),mapping=aes(y= reorder(votante_votato,count),x=count))+
  geom_col(fill="#00bfc4")+
  labs(title="Voto massimo assegnato da ogni stato",
        x="Quantità 12 punti assegnati", y= "Donatore-Ricevente")+
  geom_text(aes(label = count),nudge_y = 0.1 , nudge_x = -0.5, color="white")
```



Interessante notare come Grecia e Cipro si siano assegnati a vicenda 12 punti quasi costantemente  
 Ormai è diventata battuta ricorrente i 12 punti di Cipro e Grecia assegnati sempre l'uno all'altro ogni edizione

### Perche cipro ha ricevuto meno voti da parte della grecia?

Vediamo quante volte hanno partecipato rispettivamente Cipro e Grecia

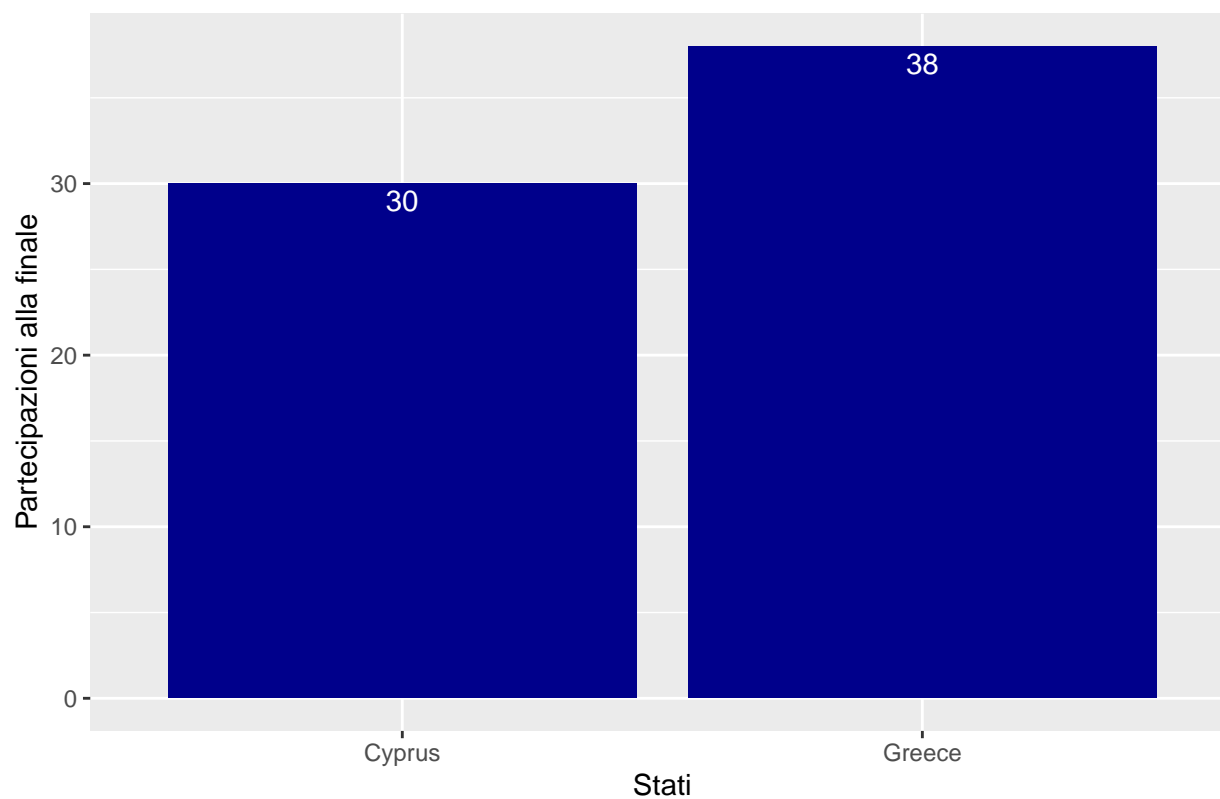
```
greece_cyprus<-events_data%>%filter(event.type=="final" | event.type=="grand-final",Country=="Greece" |
  select(Country)%>%
  arrange(Country)%>%
  group_by(Country)%>%summarise(count=n())
```

```
greece_cyprus
```

```
## # A tibble: 2 x 2
##   Country count
##   <chr>   <int>
## 1 Cyprus     30
## 2 Greece     38
```

```
ggplot(data=greece_cyprus ,mapping=aes(x=Country, y=count))+
  geom_col(fill="dark blue")+
  labs(title="Partecipazione alla finale di Cipro e Grecia",
        x="Stati", y= "Partecipazioni alla finale")+
  geom_text(aes(label = count),nudge_y = -1 , color="white")
```

Partecipazione alla finale di Cipro e Grecia



Cipro si è qualificato per le finali meno volte della grecia e, poiché uno stato partecipante può ancora votare, da il suo voto alla grecia mentre, la giuria greca, non può fare altrettanto in quanto cipro non è più parte del concorso.

## Analisi di correlazione tra punti ottenuti e ordine di apparizione

è ora di passare a qualcosa di più serio, un'analisi sulla correttezza dei voti emessi

Spesso sorge spontaneo chiedersi quanto i voti vengano influenzati da fattori esterni e spesso ciò accade dopo che il proprio paese non ha guadagnato l'ambito trofeo

Ma perché analizzare l'ordine di apparizione? Un utente sul social network Reddit ha condiviso questa domanda con la sua opinione:

- Il voto viene influenzato poiché molti telespettatori assistono solo alle esibizioni successive a quella del proprio rappresentante e non a quelle precedenti. Lo spettatore quindi si fa un'idea solo sulle esibizioni che avvengono dopo quella della propria nazione e tende a votarle dando un punteggio maggiore ad esse rispetto alle precedenti.

Ipotesi legittima che adesso andremo a verificare

### L'ordine di esibizione influenza i punti ottenuti totali?

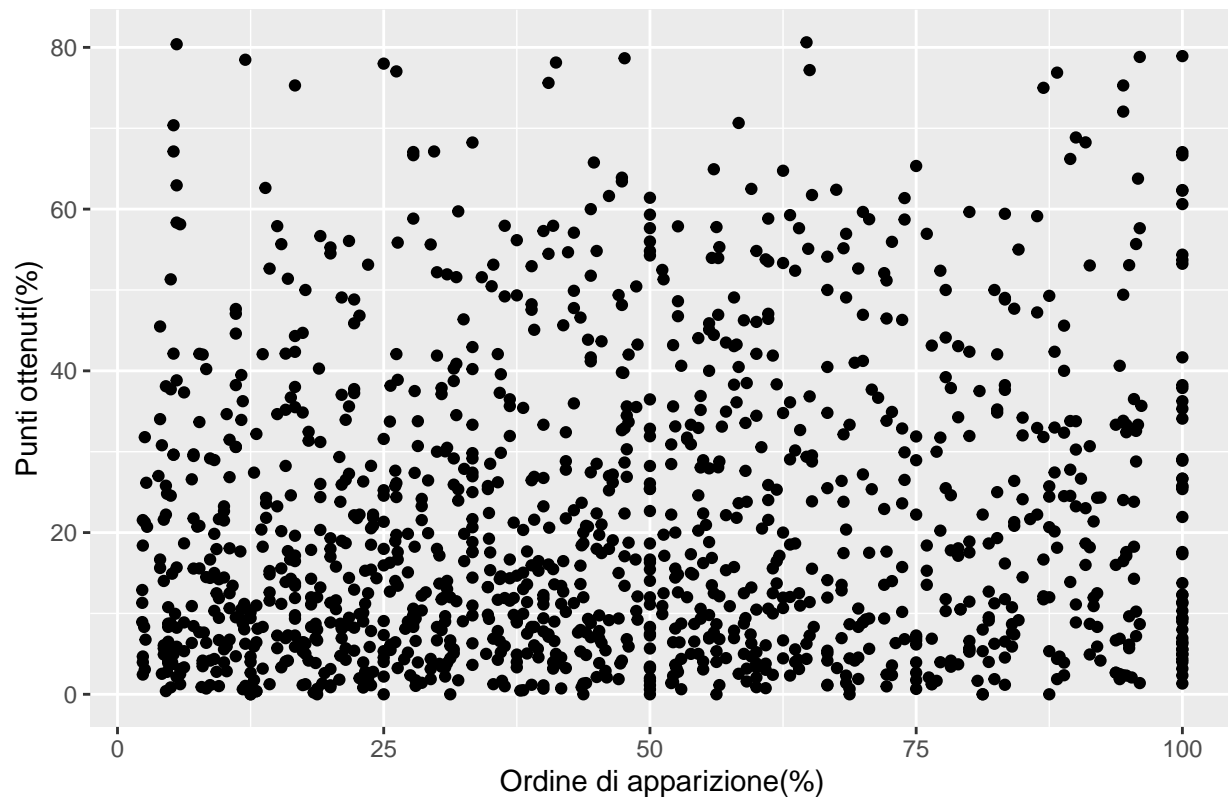
```
global_votes<-events_data%>%filter(event.type=="final" | event.type=="grand-final",event.year>1956)%>%s
support=events_data%>%filter(event.year>1956)%>%group_by(event.year)%>%summarize(partecipanti=n_distinc
global_votes=inner_join(x=global_votes ,y=support)%>%mutate(percent.appearance=100*order.of.appearance/j

## Joining, by = "event.year"
global_votes=inner_join(global_votes,edition_max_votes)%>%mutate(percent.pts=100*pts/(max_points*(partec

## Joining, by = "event.year"
ggplot(global_votes,aes(percent.appearance,percent.pts))+
  geom_point()+
  labs(title="Analisi globale",x="Ordine di apparizione(%)", y= "Punti ottenuti(%)")

## Warning: Removed 24 rows containing missing values (geom_point).
```

## Analisi globale



Da una prima occhiata non sembra esserci una correlazione tra voti e ordine di apparizione, ma per esserne sicuri guardiamo i dati matematici

```
mod1 = lm(percent.pts ~ percent.appearance, data = global_votes)
summary(mod1)
```

```
##
## Call:
## lm(formula = percent.pts ~ percent.appearance, data = global_votes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.626 -14.277  -5.332  10.376  62.040
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   17.92685    0.99497   18.02 < 0.0000000000000002 ***
## percent.appearance  0.07656    0.01832    4.18   0.0000312 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.05 on 1256 degrees of freedom
## (24 osservazioni eliminate a causa di valori mancanti)
## Multiple R-squared:  0.01372,    Adjusted R-squared:  0.01293
## F-statistic: 17.47 on 1 and 1256 DF,  p-value: 0.00003118
```

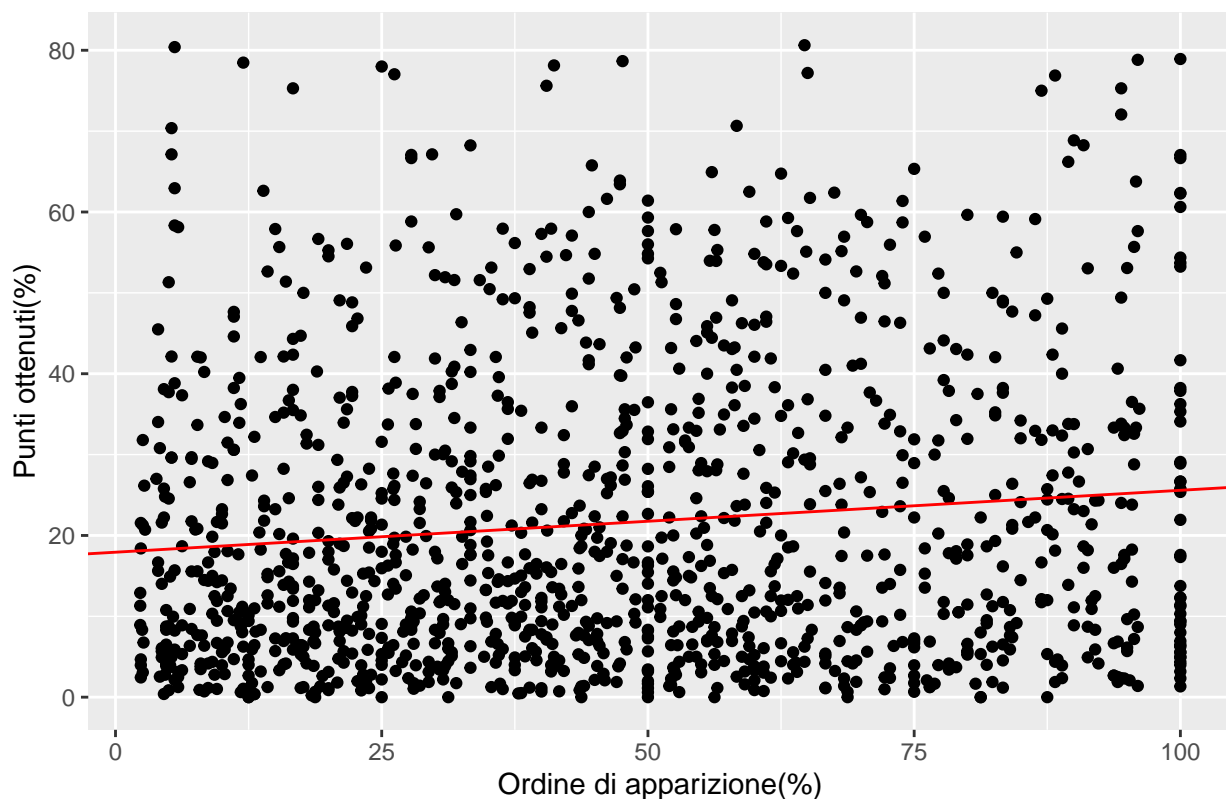
```
mod1$coefficients
```

```
##      (Intercept) percent.appearance  
##      17.92684731      0.07655813
```

```
ggplot(global_votes, aes(percent.appearance, percent.pts)) +  
  geom_point() +  
  geom_abline(intercept = mod1$coefficients[1],  
             slope = mod1$coefficients[2],  
             color = "red")+  
  labs(title="Analisi globale", x="Ordine di apparizione(%)", y= "Punti ottenuti(%)")
```

```
## Warning: Removed 24 rows containing missing values (geom_point).
```

### Analisi globale



La retta che mostra la previsione tende ad aumentare leggermente, ma non possiamo parlare ancora di correlazione

Cerchiamo ora R e R<sup>2</sup>

```
r_global = cor(global_votes$percent.appearance, global_votes$percent.pts, use= "complete.obs")  
r_global
```

```
## [1] 0.1171318
```

```
r_global^2
```

```
## [1] 0.01371985
```

Poiché r<sup>2</sup> è molto vicino a 0, una correlazione tra l'ordine di apparizione e i punti ottenuti sembra quindi non esserci.

Considero le due variabili indipendenti

Ma varrà la stessa cosa per il voto degli spettatori?

ripetiamo lo stesso procedimento ma sta volta usando i dati del televoto

```
#punti_televoto-----
tele_votes<-event_televote%>%
  filter(event.type=="grand-final")%>%
  group_by(event.year,Receiving.country)%>%
  summarise(pts=sum(Televoting.points,na.rm = TRUE))

## `summarise()` has grouped output by 'event.year'. You can override using the
## `.groups` argument.

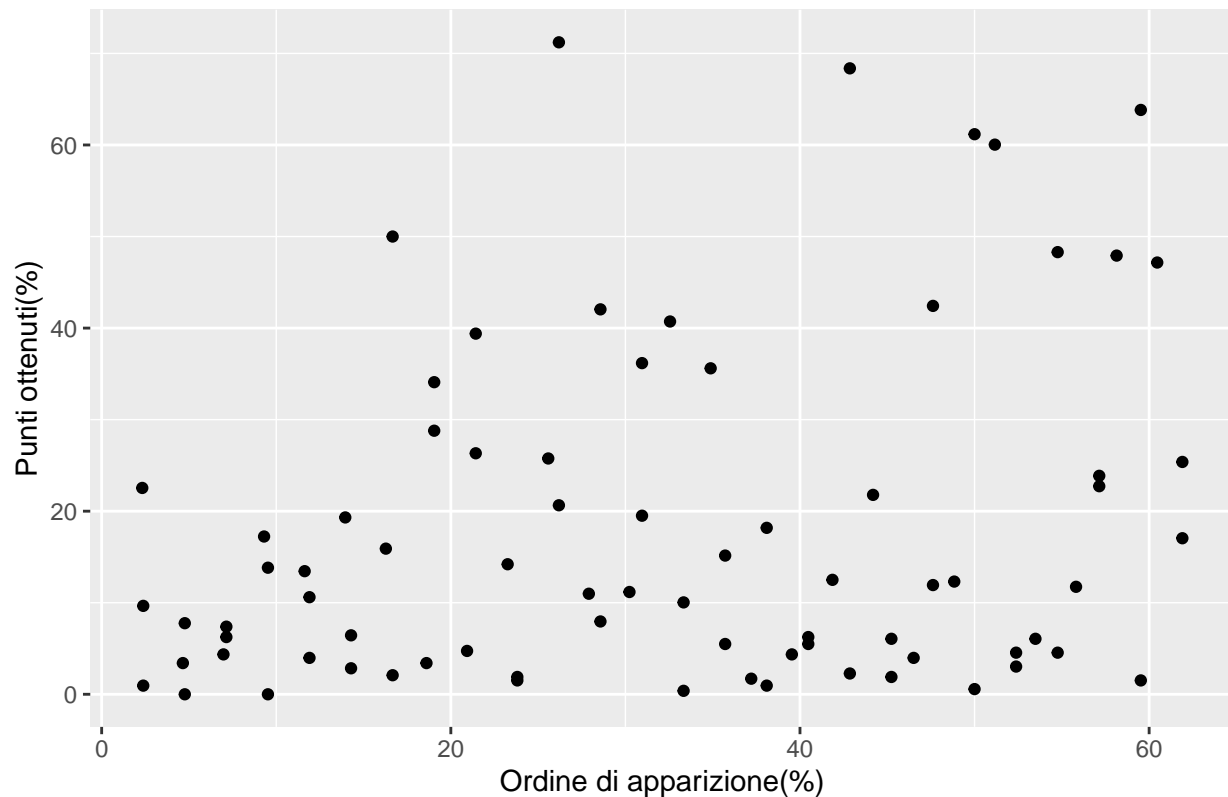
tele_votes=inner_join(x=tele_votes , y=global_votes, by=c("event.year","Receiving.country"))

tele_votes=inner_join(x=tele_votes ,y=support)%>%
  mutate(percent.appearance=100*order.of.appearance/partecipanti, percent.voti=100*pts.x/(12*nrow(event.

## Joining, by = c("event.year", "partecipanti")

ggplot(tele_votes,aes(percent.appearance,percent.voti))+geom_point()+
  labs(title="Analisi solo televoto", x="Ordine di apparizione(%)", y= "Punti ottenuti(%)")
```

Analisi solo televoto



Il grafico sembra già mostrare un certo andamento. Più ci si sposta a destra più la percentuale dei voti sale

```
mod2 = lm(percent.voti ~ percent.appearance, data = tele_votes)
summary(mod2)
```



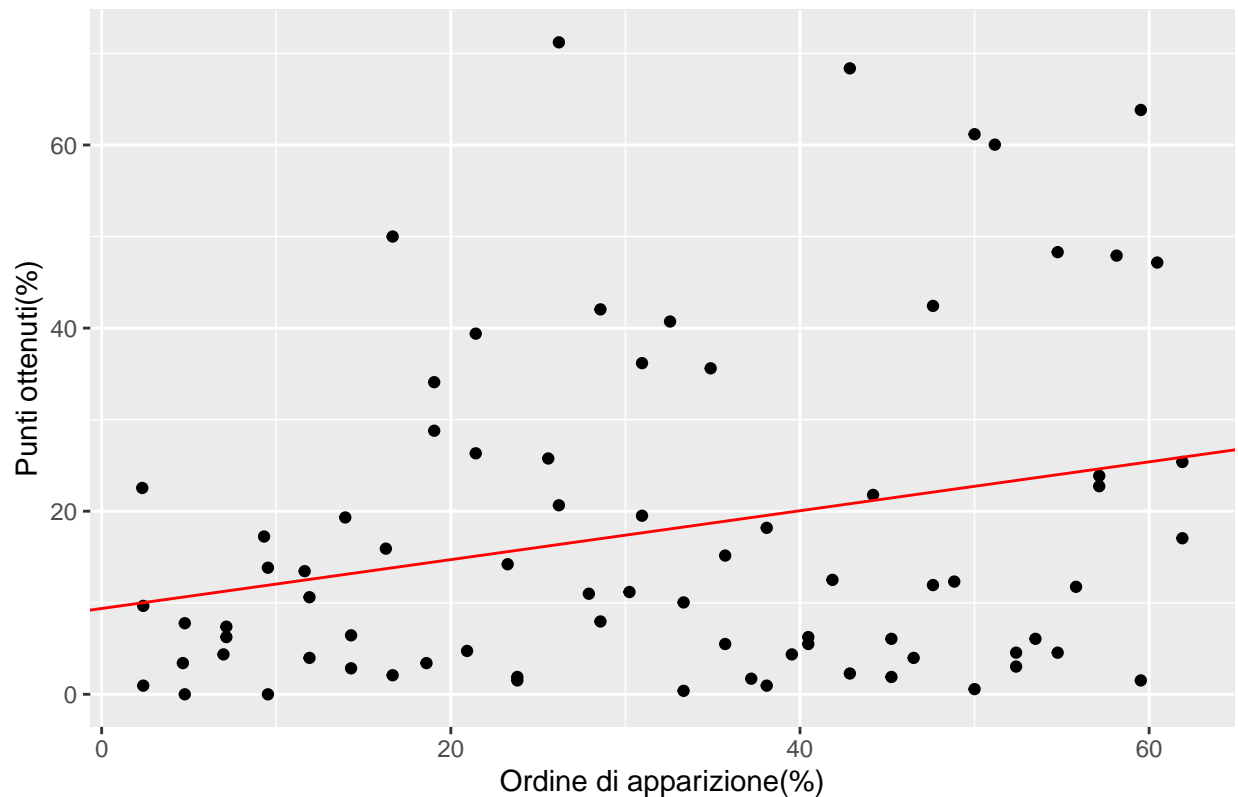
```
##
## Call:
## lm(formula = percent.voti ~ percent.appearance, data = tele_votes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.746 -12.375  -5.433   8.723  54.849
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.3720     4.1563   2.255  0.0270 *
## percent.appearance  0.2669     0.1139   2.343  0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.83 on 76 degrees of freedom
## Multiple R-squared:  0.06739,    Adjusted R-squared:  0.05512
## F-statistic: 5.492 on 1 and 76 DF,  p-value: 0.02173
```

```
mod2$coefficients
```

```
##      (Intercept) percent.appearance
##      9.371951      0.266938
```

```
ggplot(tele_votes, aes(percent.appearance,percent.voti)) +
  geom_point() +
  geom_abline(intercept = mod2$coefficients[1],
              slope = mod2$coefficients[2],
              color = "red")+
  labs(title="Analisi solo televoto",x="Ordine di apparizione(%)", y= "Punti ottenuti(%)")
```

## Analisi solo televoto



La retta ci mostra quello che sospettavamo ma anche in questo caso è un aumento lieve

```
r_tele = cor(tele_votes$percent.appearance, tele_votes$percent.voti, use= "complete.obs")
r_tele
```

```
## [1] 0.2595921
```

```
r_tele^2
```

```
## [1] 0.06738805
```

Possiamo vedere che, sebbene il televoto sia leggermente più correlato all'ordine di apparizione rispetto al voto globale, siamo ancora lontani dal considerare voti e ordine di apparizione correlati.

### Vale anche per i singoli stati?

Valutiamo la correlazione dei voti emessi dalle giurie dei singoli stati e l'ordine di comparsa dello stato votato

Uniamo ora i nostri dati con le tabelle di supporto al fine di ottenere una tabella utile ai nostri scopi

```
state_votes=inner_join(state_votes,edition_max_votes)
```

```
## Joining, by = "event.year"
```

```
temp=events_data%>%filter(event.type=="final" | event.type=="grand-final",event.year>1956)%>%
  select(event.year,Country,order.of.appearance)
```

```
state_votes=inner_join(state_votes,temp,by=c("event.year"="event.year","recipient.country"="Country"))%>%
```

```
state_votes
```

```
## # A tibble: 14,031 x 7
##   event.year awarding.country recipient.country pts max_points
##   <dbl> <chr> <chr> <dbl> <dbl>
## 1 1957 Austria The Netherlands 6 10
## 2 1957 Austria Luxembourg 3 10
## 3 1957 Austria United Kingdom 1 10
## 4 1957 Belgium The Netherlands 5 10
## 5 1957 Belgium France 2 10
## 6 1957 Belgium United Kingdom 1 10
## 7 1957 Belgium Italy 1 10
## 8 1957 Belgium Germany 1 10
## 9 1957 Denmark The Netherlands 3 10
## 10 1957 Denmark Belgium 2 10
## # ... with 14,021 more rows, and 2 more variables: order.of.appearance <dbl>,
## # pts.percent <dbl>
```

Ora raggruppiamo i dati in base allo stato votante così da avere i dati che ci servono su un'unica riga

```
grouped_states=state_votes%>%group_by(awarding.country)%>%nest()
grouped_states[1]
```

```
## # A tibble: 52 x 1
## # Groups:   awarding.country [52]
##   awarding.country
##   <chr>
## 1 Austria
## 2 Belgium
## 3 Denmark
## 4 France
## 5 Germany
## 6 Italy
## 7 Luxembourg
## 8 Switzerland
## 9 The Netherlands
## 10 United Kingdom
## # ... with 42 more rows
```

```
grouped_states[[2]][1]
```

```
## [[1]]
## # A tibble: 443 x 6
##   event.year recipient.country pts max_points order.of.appearance pts.percent
##   <dbl> <chr> <dbl> <dbl> <dbl> <dbl>
## 1 1957 The Netherlands 6 10 6 60
## 2 1957 Luxembourg 3 10 2 30
## 3 1957 United Kingdom 1 10 3 10
## 4 1958 France 7 10 3 70
## 5 1958 Germany 1 10 8 10
## 6 1958 Italy 1 10 1 10
## 7 1958 Sweden 1 10 5 10
## 8 1959 The Netherlands 3 10 5 30
## 9 1959 Denmark 2 10 2 20
## 10 1959 United Kingdom 2 10 10 20
## # ... with 433 more rows
```

Ora che abbiamo reso il nostro dataframe più ordinato possiamo agire su di esso e calcolare la correlazione

tra voti e apparizione per ogni singolo stato con la funzione `map()`

```
states_model=function(df)
{
  lm(order.of.appearance ~ pts.percent, data=df)
}

grouped_states=grouped_states%>%mutate(model=map(data,states_model))

grouped_states
```

```
## # A tibble: 52 x 3
## # Groups:   awarding.country [52]
##   awarding.country data          model
##   <chr>          <list>          <list>
## 1 Austria      <tibble [443 x 6]> <lm>
## 2 Belgium      <tibble [521 x 6]> <lm>
## 3 Denmark      <tibble [399 x 6]> <lm>
## 4 France       <tibble [525 x 6]> <lm>
## 5 Germany      <tibble [529 x 6]> <lm>
## 6 Italy         <tibble [353 x 6]> <lm>
## 7 Luxembourg   <tibble [293 x 6]> <lm>
## 8 Switzerland  <tibble [501 x 6]> <lm>
## 9 The Netherlands <tibble [508 x 6]> <lm>
## 10 United Kingdom <tibble [543 x 6]> <lm>
## # ... with 42 more rows
```

Decomprimiamo la tabella ed estrapoliamo  $R^2$  per verificare se esistono stati influenzati nel loro voto

```
grouped_states=grouped_states %>% mutate(data = map2(data, model, add_residuals))

grouped_states = grouped_states %>% mutate(glance = map(model, broom::glance))

states_dependence=unnest(grouped_states,glance)
states_dependence=states_dependence%>%arrange(-r.squared)%>%select(awarding.country,r.squared)

states_dependence
```

```
## # A tibble: 52 x 2
## # Groups:   awarding.country [52]
##   awarding.country r.squared
##   <chr>          <dbl>
## 1 Czech Republic    0.0463
## 2 Morocco           0.0441
## 3 Poland            0.0290
## 4 Azerbaijan        0.0259
## 5 Yugoslavia        0.0230
## 6 Estonia           0.0226
## 7 The Netherlands   0.0174
## 8 Ireland           0.0169
## 9 Bulgaria          0.0161
## 10 United Kingdom    0.0157
## # ... with 42 more rows
```

Come visto in precedenza siamo ancora lontani dal definire una correlazione tra voti e ordine di apparizione. Possiamo però notare che Marocco e Repubblica Ceca hanno valori nettamente più alti rispetto agli altri stati.

Cio è dovuto al basso numero di partecipazioni(visibile nelle tabelle mostrate all'inizio). Hanno rispettivamente partecipato 1 e 3 volte alla finale degli eurovision quindi non si hanno sufficienti dati per esprimere una correlazione.

In conclusione possiamo dire che l'ordine di apparizione non influenza gli esiti finali, quindi il sistema di voto attuale giusto e imparziale.