

# Són la positivitat, l'energia i la dansabilitat factors compartits entre les cançons més escoltades a Finlàndia i Espanya?

Xavier Coll Ribas - Enric Condal Asensio - Roc Salvador Andreazini

Desembre 2020

## 1. RESUM

**Objectiu:** Determinar si, fent cas als estereotips, realment són més positives, energètiques i ballables les cançons més escoltades a Espanya que les més escoltades a Finlàndia.

**Mètodes:** Obtenció de dades de les 50 cançons més populars d'ambdós països mitjançant l'API de Spotify i un codi en Python, i extracció de conclusions mitjançant l'ús de procediments estadístics.

**Resultats:** Els valors de la valència, energia i dansabilitat en les cançons més escoltades a Espanya són més alts que els obtinguts pels mateixos atributs en les cançons més escoltades a Finlàndia.

**Discussió:** Després d'analitzar les dades i amb l'ajut de proves d'hipòtesis unilaterals, hem arribat a la conclusió que els valors de positivitat, energia i dansabilitat no són iguals en els dos països i que tenim certa evidència de que les cançons més escoltades a Espanya, els valors són més grans que en les cançons més escoltades a Finlàndia.

## 2. INTRODUCCIÓ

### Justificació

Per una banda, el tema de la música és una qüestió en la qual moltes persones tenen una opinió ben formada i els hi és proper el fet de les diferents emocions que dues cançons poden arribar a generar. Tanmateix, per una altra banda, els estereotips socials entre països són una classe de prejudicis que, més fidels o menys a la realitat, continuen presents a la societat actual del segle XXI.

Per aquesta raó, el nostre projecte pot esdevenir una eina útil per tal d'arribar a conèixer més a fons les característiques socials que es deriven dels gustos musicals dels habitants d'Espanya i Finlàndia: dos països amb diferències geogràfiques i culturals òbvies, però que gran part de la societat justifica les diferències socials amb la idea dels mateixos estereotips, sense qüestionar la veracitat d'aquests.

## Objectius

La música. Un art basat en la combinació i manipulació de sons, que podria arribar a semblar pla i senzill per aquell qui en desconeix les possibilitats, però que sense cap mena de dubte, s'ha convertit avui en dia en un dels principals motors de la societat.

I realment, cal admetre que la comparativa de la seva força motora no ha estat resultat de l'atzar. Qui podria discutir que no ha seguit mai amb el peu la pulsació d'aquella música que sonava de fons? O que no ha canviat la seva perspectiva després de fer una travessa per la seva cançó preferida? O que no s'ha sentit motivat pel ritme d'una melodia mentre feia esport? En la societat actual, els efectes que pot tenir una cançó concreta en un instant determinat són innegables. Al més pur estil d'estratègia de màrqueting: una cançó pot ser vertaderament influent.

Però la cosa no es queda aquí. Arreu del món és coneguda la influència de la música. Però aquesta influència, evoca les mateixes emocions en una societat i en una altra? Dos habitants separats per tres mil quilòmetres de terra i mar, busquen el mateix tipus de sentiment en les cançons que més escolten?

I és amb la idea que genera aquesta pregunta, a partir de la qual nosaltres hem decidit desenvolupar el nostre projecte. Mantenint l'objectiu en recollir dades i comparar l'energia, la dansabilitat i la valència (o positivitat) de les 50 cançons més populars de Finlàndia i Espanya, dos països de la Unió Europea separats per més de tres mil quilòmetres i amb uns estereotips socials ben definits per les seves cultures, amb fama de més energètiques i positives en el cas d'Espanya i més calmades i tranquil·les en el cas de Finlàndia.

## 3. PREMISSES I HIPÒTESI

Fent referència a la pregunta:

*“Són la positivitat, l'energia i la dansabilitat factors compartits entre les cançons més escoltades a Finlàndia i Espanya?”*

I establint les següents premisses (en les que s'aprofundirà en l'apartat següent de recollida de dades i disseny de l'estudi):

- A cada cançó li corresponen tres valors naturals de 0 a 1, que indiquen la seva positivitat, energia i dansabilitat amb 3 posicions decimals.
- Un valor proper a 1 representa un grau elevat del factor que s'analitza; i un valor proper a 0, un grau baix (exemple: en termes de la positivitat, una cançó amb un valor de 0.896 és més positiva que una cançó amb un valor de 0.345)

Es considerarà la següent hipòtesi:

*“Les cançons més escoltades a Espanya tenen valors de positivitat, energia i dansabilitat **més alts** que les cançons més escoltades a Finlàndia.”*

## 4. RECOLLIDA DE DADES I DISSENY DE L'ESTUDI

Les dades es recolliran mitjançant un petit programa de Python, el qual fa ús de la llibreria Spotipy<sup>1</sup> que permet obtenir dades de totes les cançons disponibles a la plataforma.

El programa consisteix en:

1. Identificació de l'usuari mitjançant la creació de les credencials d'una aplicació ([My Dashboard](#)).
2. Input de l'URI de la playlist (es pot obtenir fàcilment a l'aplicació de Spotify, anant al botó de share i copiant l'URI) de la que es volen treure les dades.
3. Obtenció de la ID de la playlist del Top 50 a nivell mundial i les dades de cada cançó.
4. Guardat de les dades en un fitxer CSV

Les dades seran recollides el mateix dia en què es comenci l'anàlisi de les dades, ja que el procés d'obtenció d'aquestes és molt ràpid. Aclarem també que el Top 50 de Spotify s'actualitza cada dilluns<sup>2</sup>.

### Pla estadístic

- Per tal d'extreure (o no) conclusions de les dades que obtindrem, es duran a terme tres proves d'hipòtesis sobre el paràmetre  $\mu$ , on en cada cas correspondrà a l'esperança de la valència, l'energia o la dansabilitat.
- Tenint en ment la hipòtesis plantejada, les proves seran unilaterals
- Les mostres aleatòries simples son independents
- Els valors dels atributs de les cançons de cada país presenten normalitat

## 5. VARIABLES

Les variables recollides en la nostra mostra són les següents:

- **Positivitat/València:** Sentiment de positivitat que transmet una cançó. Com més negativa sigui una cançó, menys puntuació rebrà. Aquesta variable pren valors del 0.000 al 1.000.
- **Energia:** Es mesura segons la intensitat i l'activitat d'una cançó. Cançons, com ara, del gènere del Metal, obtindran molta puntuació gràcies a les guitarres, la veu i la bateria, en canvi, una cançó d'orquestra, al ser més calmada, rebrà una puntuació més baixa. Els seus valors ocupen posicions dins un rang de 0.000 a 1.000.
- **Dansabilitat:** Quant de 'ballable' és una cançó, basant-se en el seu ritme, tempo i potència. Pren valors de 0.000 a 1.000.
- **País/Procedència:** En el nostre projecte, és una variable categòrica binària, on pot prendre el valor "Espanya" o bé "Finlàndia", depenent del país al que es refereixen les dades de una cançó determinada.

---

<sup>1</sup> Welcome to Spotipy!¶. (n.d.). Retrieved from <https://spotipy.readthedocs.io/en/2.16.1/>

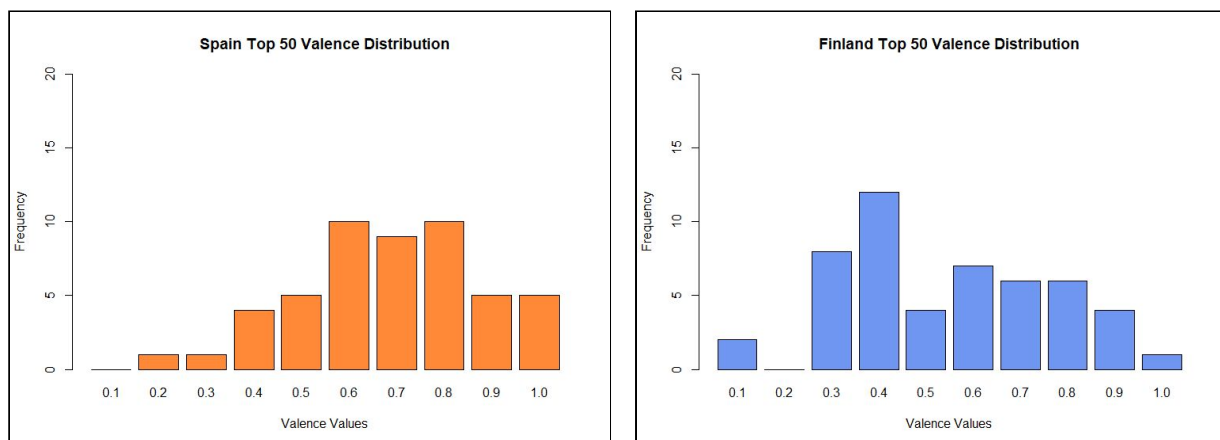
<sup>2</sup> Bernardo, Á. (2013, May 22). El Top 50 de Spotify: Sus canciones más escuchadas y compartidas. Retrieved from <https://hipertextual.com/2013/05/spotify-mostrara-su-top-50>

## 6. RESULTATS

Usant el mètode descrit anteriorment, en data **11 de desembre de 2020**, hem obtingut valors de valència, energia i dansabilitat per cada una de les 50 cançons més escoltades a Espanya i a Finlàndia (50 cançons de cada país).

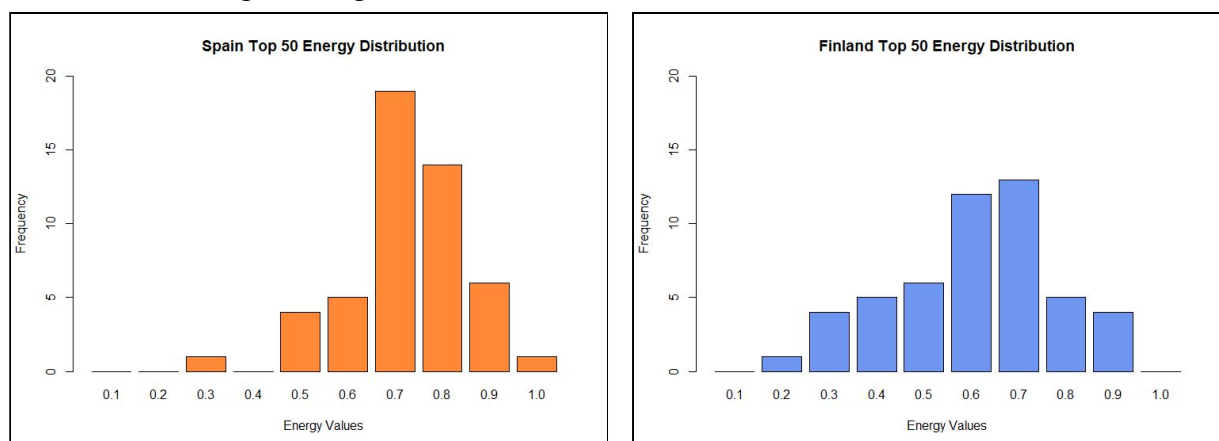
Per tal de fer-ne una visualització més eficient, hem representat la freqüència absoluta de cançons segons un dels atributs en cada gràfics de barres.

Els valors de la positivitats obtinguts són:



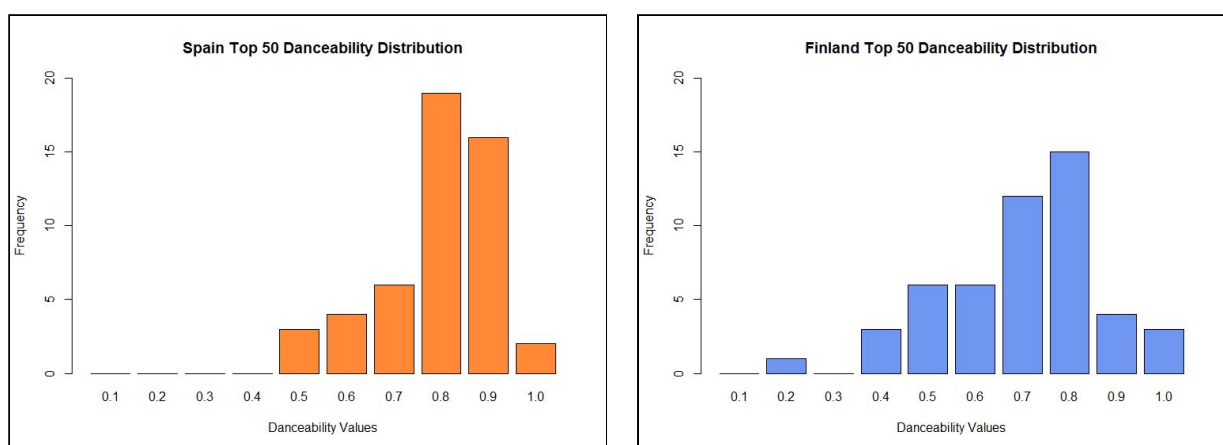
**Imatges 1 i 2. Distribució dels valors de la positivitats/valència de les 50 cançons més escoltades a Espanya (esquerra, en taronja) i Finlàndia (dreta, en blau).**

Els valors de l'energia obtinguts són:



**Imatges 3 i 4. Distribució dels valors de l'energia de les 50 cançons més escoltades a Espanya (esquerra) i Finlàndia (dreta).**

Els valors de la dansabilitat obtinguts són:



**Imatges 5 i 6. Distribució dels valors de la dansabilitat de les 50 cançons més escoltades a Espanya (esquerra) i Finlàndia (dreta).**

## 7. ANÀLISIS I DISCUSSIÓ

Primerament, farem una síntesi de de les mesures estadístiques que ens poden ser útils per entendre els resultats obtinguts.

En el cas del Top 50 d'Espanya:

Atribut	Min	Quantil 0,25	Mitjana	Mediana	Quantil 0,75	Max	Desviació tipus mostral
València	0.1450	0.5137	0.6421	0.6785	0.7705	0.965	0.1903
Energia	0.2380	0.5865	0.6749	0.6730	0.7635	0.922	0.1311
Dansabilitat	0.4460	0.6691	0.7454	0.7670	0.8215	0.917	0.1128

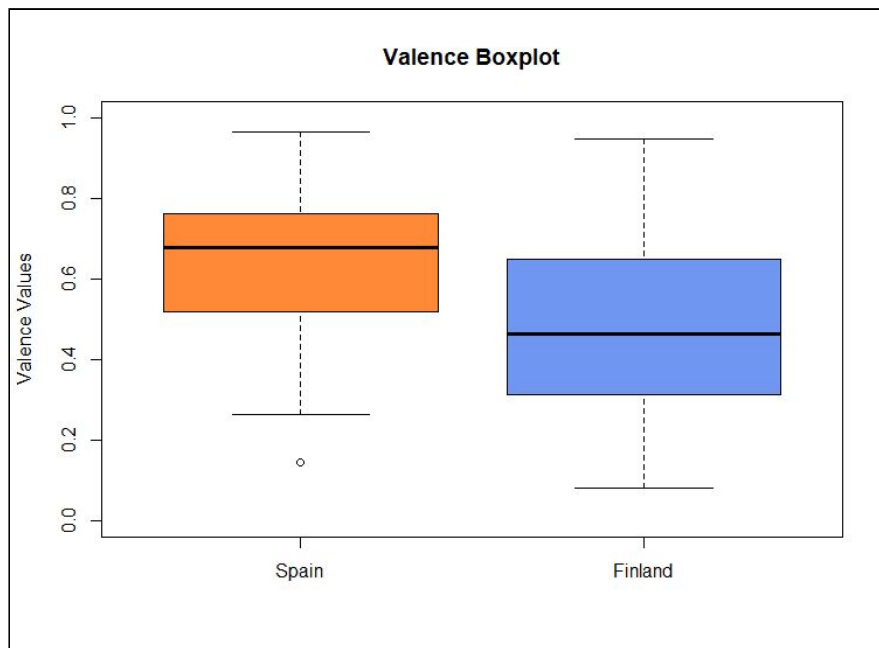
**Taula 1. Estadístics de la mostra de les 50 cançons d'Espanya.**

En el cas del Top 50 de Finlàndia:

Atribut	Min	Quantil 0,25	Mitjana	Mediana	Quantil 0,75	Max	Desviació tipus mostral
València	0.0799	0.3410	0.4884	0.4645	0.6358	0.9470	0.2186
Energia	0.1840	0.4353	0.5516	0.5845	0.6678	0.8920	0.1724
dansabilitat	0.1750	0.5402	0.6498	0.6710	0.7595	0.9350	0.1625

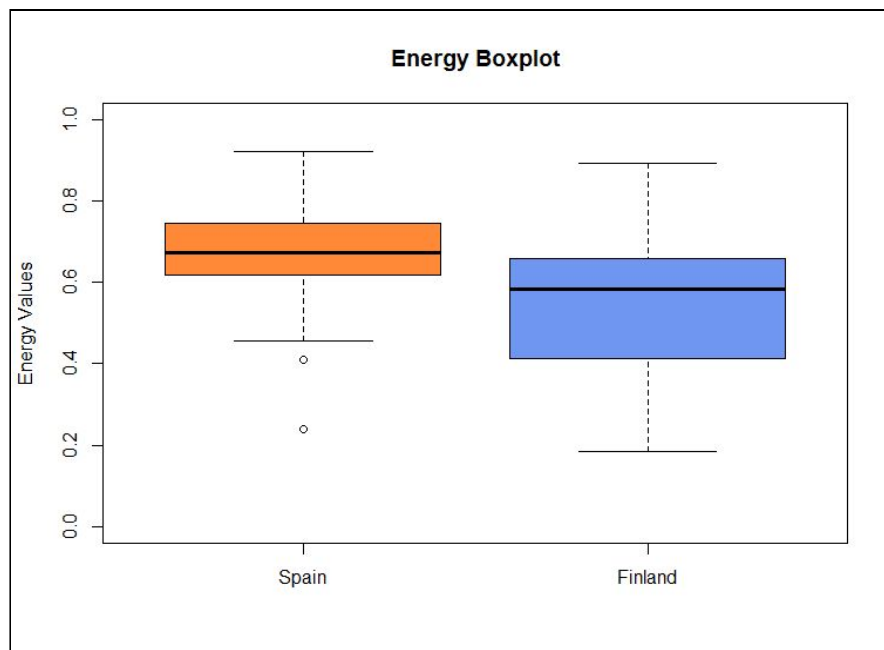
**Taula 2. Estadístics de la mostra de les 50 cançons de Finlàndia.**

A més, és bona idea graficar tres boxplots diferents (un per cada atribut) per tal d'interpretar d'una manera més visual el significat de les dades.



**Imatge 7. Boxplot de la valència de les 50 cançons més escoltades a Espanya (esquerra, en taronja) i a Finlàndia (dreta, en blau).**

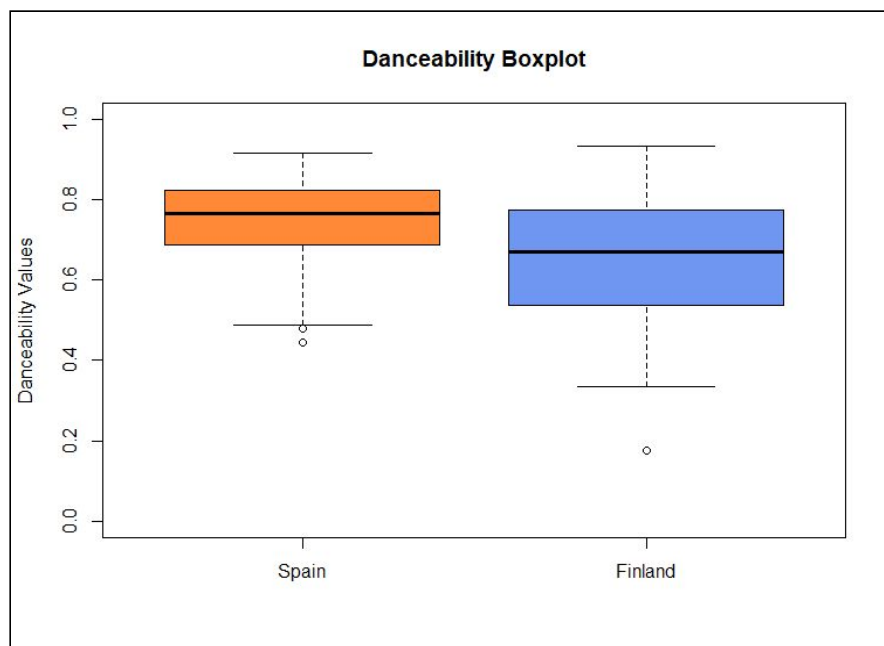
En el **boxplot de la valència/positivitat**, es veu com la mediana de les cançons més escoltades a Espanya és més elevada, així com també ho és la caixa del rang interquartil. S'aprecia també en el cas d'Espanya, un únic cas atípic, molt allunyat dels altres (on valència = 0.145). La dispersió en ambdós països és elevada, tot i que a Finlàndia, l'atribut de la valència pren valors dins d'un rang més gran.



**Imatge 8. Boxplot de l'energia de les 50 cançons més escoltades a Espanya (esquerra) i a Finlàndia (dreta).**

En el **boxplot de l'energia**, les medianes són properes, tot i que el rang interquartílic d'Espanya és molt més reduït que el de Finlàndia. En aquest cas, la diferència en el rang de

valors que pren l'energia és més notable que en l'anterior gràfic i els dos casos de cançons amb valors atípicament baixos a Espanya es troben fins i tot dins del rang inferior del 25% de les cançons de Finlàndia.



**Imatge 9. Boxplot de la dansabilitat de les 50 cançons més escoltades a Espanya (esquerra) i a Finlàndia (dreta).**

En el **boxplot de la dansabilitat**, les medianes són similars i la caixa del rang interquartílic de Finlàndia és més gran que la d'Espanya. En aquest cas, ambdues gràfiques presenten valors més alts i el rang superior del 25% de Finlàndia arriba a valors més alts que el mateix rang d'Espanya. A més, els dos valors atípics d'Espanya són propers al límit inferior del bigoti, mentre que el valor atípic de Finlàndia es troba més allunyat que el límit inferior del bigoti.

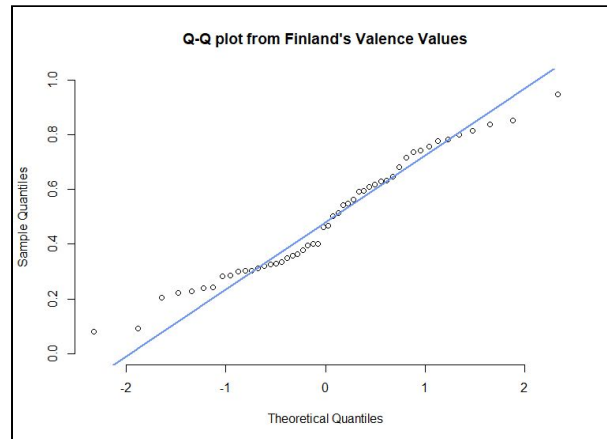
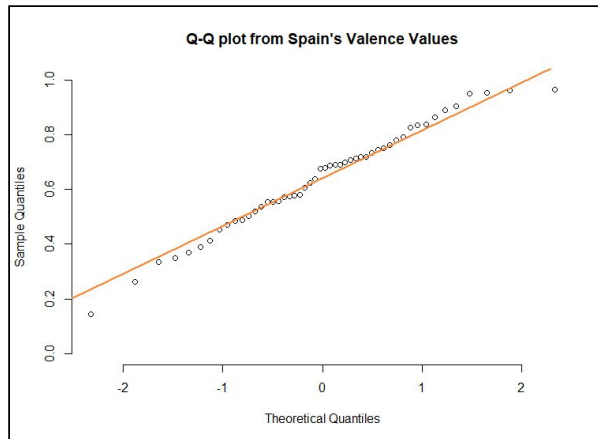
A continuació, amb l'objectiu de determinar si realment tenim suficient evidència a partir de les mostres obtingudes per concloure que els valors de la positivitat, l'energia i la dansabilitat són més alts a les cançons més escoltades a Espanya que les més escoltades a Finlàndia, hem realitzat tres **proves d'hipòtesis** per cada un dels atributs.

### **Proves d'hipòtesis en 2 mostres:**

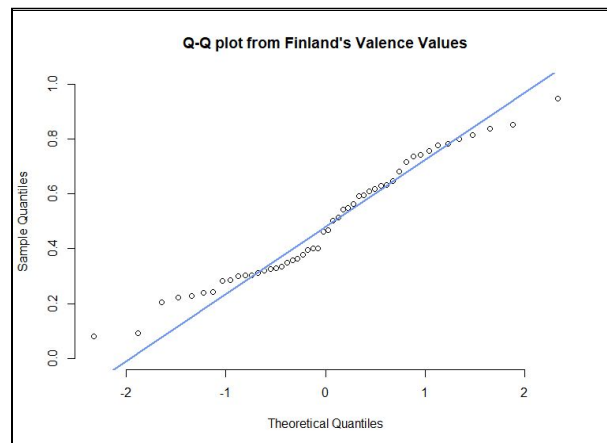
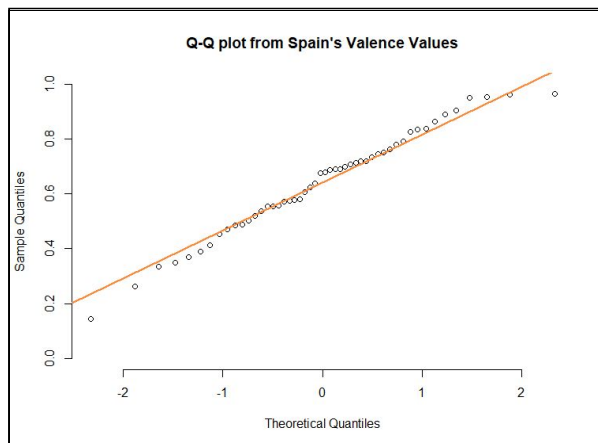
En les tres proves d'hipòtesis:

- Les mostres són aleatòries simples i independents entre elles (premissa d'independència)
- Les tres variables que mesuren les característiques de les cançons de cada país, observant els gràfics Q-Q presentats a continuació, podem dir que segueixen una distribució Normal (premissa de normalitat)

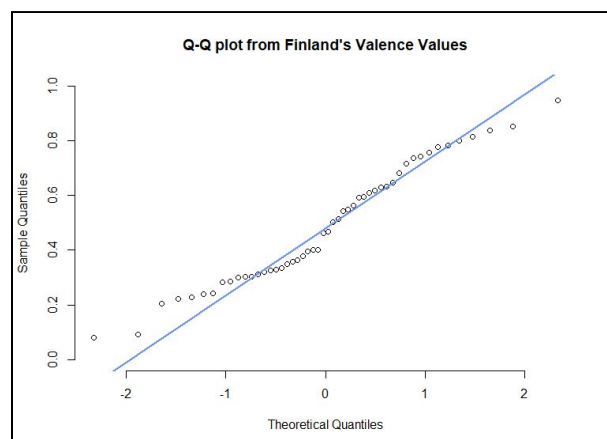
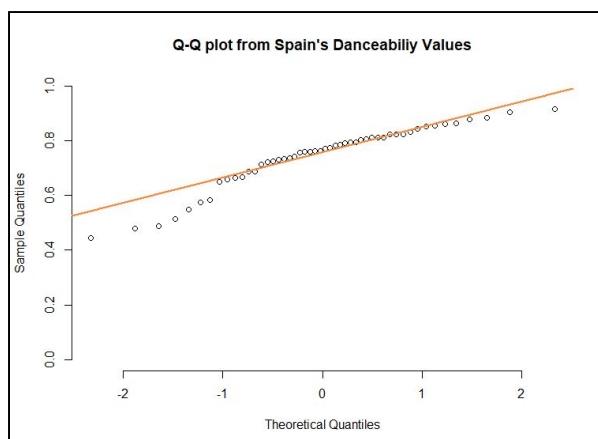
- Assumim que les variàncies de les dues poblacions comparades son iguals i desconegudes (premissa d'igualtat de variància o homoscedasticitat)
- Es treballarà amb un risc de  $\alpha = 0.05$



**Imatges 10 i 11. Gràfics Q-Q de la valència de les 50 cançons més escoltades a Espanya (esquerra, en taronja) i a Finlàndia (dreta, en blau).**



**Imatges 12 i 13. Gràfics Q-Q de l'energia de les 50 cançons més escoltades a Espanya (esquerra, en taronja) i a Finlàndia (dreta, en blau).**



**Imatges 14 i 15. Gràfics Q-Q de la dansabilitat de les 50 cançons més escoltades a Espanya (esquerra, en taronja) i a Finlàndia (dreta, en blau).**



## Comparació de les mitjanes de la valència

1. **Variables:**  $V_f$  = valència d'una cançó a Finlàndia  
 $V_s$  = valència d'una cançó a Espanya

2. **Hipòtesis:**  $H_0: \mu_{Vf} = \mu_{Vs}$   
 $H_1: \mu_{Vf} < \mu_{Vs}$

3. **Estadístic:**

$$T = \frac{\bar{V}_f - \bar{V}_s}{S * \sqrt{1/n_f + 1/n_s}}$$
$$S^2 = \frac{(n_f - 1) * S_f^2 + (n_s - 1) * S_s^2}{n_f + n_s - 2}$$

4. **Distribució estadística:**  $t \rightarrow t_{nf+ns-2} = t_{98}$

5. **Càlculs:**  $S_v = 0.1903$   
 $S_f = 0.2186$   
 $\bar{V}_s = 0.6421$   
 $\bar{V}_f = 0.4884$   
 $n_f = n_s = 50$

Amb aquestes dades podem calcular l'estadístic:

$$S^2 = ((50-1)*0.2186^2 + (50-1)*0.1903^2)/(50+50-2) = 0.042$$

$$S = \text{sqrt}(0.042) = 0.2045$$

$$T = (0.4884 - 0.6421)/(0.2045*\text{sqrt}(1/50 + 1/50)) = -3,749$$

6. **P-valor** = 0.0001499

7. **Conclusió:** Al obtenir un P-valor de 0.0001499, sent aquest bastant més petit que 0.05, rebutjem  $H_0$ .

Conclusió pràctica:

Per tant, després d'observar els boxplots i ara, obtenint els resultats teòrics, podem dir amb certa evidència que el Top 50 de Finlàndia té una valència menor a la del Top 50 espanyol.

```
Two Sample t-test

data:  spainData$valence and finlandData$valence
t = 3.7495, df = 98, p-value = 0.0001499
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.08561284      Inf
sample estimates:
mean of x mean of y
 0.642080  0.488412
```

Imatge 16. Output de R per a la prova d'hipòtesis de comparació de les mitjanes de la valència.<sup>3</sup>

<sup>3</sup> Codi utilitzat: `t.test(spainData$valence, finlandData$valence, alternative = "greater", mu = 0, paired=FALSE, var.equal=TRUE, conf.level = 0.95)`

## Comparació de les mitjanes de l'energia

1. **Variable:**  $E_f$  = energia d'una cançó a Finlàndia

$E_s$  = energia d'una cançó a Espanya

2. **Estadístic:**

$$T = \frac{\bar{E}_f - \bar{E}_s}{S \cdot \sqrt{1/n_f + 1/n_s}}$$
$$S^2 = \frac{(n_f - 1) \cdot S_f^2 + (n_s - 1) \cdot S_s^2}{n_f + n_s - 2}$$

3. **Hipòtesis:**

$$H_0: \mu_{Ef} = \mu_{Es}$$

$$H_1: \mu_{Ef} < \mu_{Es}$$

4. **Distribució estadística:**  $t \rightarrow t_{n_f+n_s-2} = t_{98}$

5. **Càlculs:**

$$S_s = 0.1311$$

$$S_f = 0.1724$$

$$E_s = 0.6749$$

$$E_f = 0.5516$$

$$n_f = n_s = 50$$

Amb aquestes dades podem calcular l'estadístic:

$$S^2 = ((50-1) \cdot 0.1724^2 + (50-1) \cdot 0.1311^2) / (50+50-2) = 0.0235$$

$$S = \sqrt{0.0235} = 0.1533$$

$$T = (0.5516 - 0.6749) / (0.1533 \cdot \sqrt{1/50 + 1/50}) = -4.021526$$

6. **P-valor** = 0.00005554

7. **Conclusió:** Un cop fet els càlculs, obtenim un P-valor de  $5.554 \cdot 10^{-5}$ . Podem veure que el p-valor obtingut és més petit que 0.05, per tant, rebutjem  $H_0$ .

Conclusió pràctica: Un cop amb els valors teòrics i les gràfiques, arribem a la conclusió que l'energia de les cançons del Top 50 de Finlàndia és menor a les del Top 50 d'Espanya.

```
Two Sample t-test

data: spainData$energy and finlandData$energy
t = 4.028, df = 98, p-value = 5.554e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.07250478      Inf
sample estimates:
mean of x mean of y
 0.67492   0.55156
```

Imatge 17. Output de R per a la prova d'hipòtesis de comparació de les mitjanes de l'energia.<sup>4</sup>

<sup>4</sup> Codi utilitzat: `t.test(spainData$energy, finlandData$energy, alternative = "greater", mu = 0, paired=FALSE, var.equal=TRUE, conf.level = 0.95)`

## Comparació de les mitjanes de la dansabilitat

1. **Variable:**  $D_f$  = dansabilitat d'una cançó a Finlàndia  
 $D_s$  = dansabilitat d'una cançó a Espanya
2. **Estadístic:**

$$T = \frac{\overline{D_f} - \overline{D_s}}{S * \sqrt{1/n_f + 1/n_s}}$$
$$S^2 = \frac{(n_f - 1) * S_f^2 + (n_s - 1) * S_s^2}{n_f + n_s - 2}$$

3. **Hipòtesis:**  $H_0: \mu_{Df} = \mu_{Ds}$   
 $H_1: \mu_{Df} < \mu_{Ds}$

4. **Distribució estadística:**  $t^{\wedge} \rightarrow t_{n_f+n_s-2} = t_{98}$

5. **Càlculs:**  $S_s = 0.1128$   
 $S_f = 0.1625$   
 $D_s = 0.7454$   
 $D_f = 0.6497$   
 $n_f = n_s = 50$

Amb aquestes dades podem calcular l'estadístic:

$$S^2 = ((50-1)*0.1625^2 + (50-1)*0.1128^2)/(50+50-2) = 0.0196$$

$$S = \text{sqrt}(0.0196) = 0.1400$$

$$T = (0.6497 - 0.7454)/(0.1400 * \text{sqrt}(1/50 + 1/50)) = -3.417857$$

6. **P-valor** = 0.0004598
7. **Conclusió:** Un cop fet els càlculs, obtenim un P-valor de 0.0004598. Podem veure que el p-valor obtingut és més petit que 0.05, per tant, rebutgem  $H_0$ .

Conclusió pràctica: Un cop amb els valors teòrics i les gràfiques, arribem a la conclusió que la dansabilitat de les cançons del Top 50 de Finlàndia és menor a les del Top 50 d'Espanya.

```
Two Sample t-test

data:  spainData$danceability and finlandData$danceability
t = 3.4182, df = 98, p-value = 0.0004598
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.04916839      Inf
sample estimates:
mean of x mean of y
 0.74538   0.64976
```

Imatge 18. Output de R per a la prova d'hipòtesis de comparació de les mitjanes de la dansabilitat.<sup>5</sup>

<sup>5</sup> Codi utilitzat: `t.test(spainData$danceability, finlandData$danceability, alternative = "greater", mu = 0, paired=FALSE, var.equal=TRUE, conf.level = 0.95)`

## **8. CONCLUSIÓ**

La resposta a la pregunta de si la positivitat, l'energia i la dansabilitat són factors que comparteixen les cançons més escoltades a Finlàndia i Espanya, amb els resultats obtinguts seria que no, ja que hem descartat per cada un dels factors que la mitjana fos igual entre els dos països. A més, al fer proves unilaterals, podem concloure que en les cançons més escoltades a Espanya els factors de la positivitat, energia i dansabilitat són més elevats que en les cançons més escoltades a Finlàndia, donant certa credibilitat i evidència al tòpic de que, en aquest cas en el tema dels gustos musicals, un país mediterrani com Espanya té tendència a ser més alegre i dinàmic que un país nòrdic com Finlàndia.

### **Limitacions**

Degut a l'època de l'any en la que es va realitzar el projecte (principis de desembre), la proximitat de la festa de Nadal ha provocat la reaparició de cançons típiques d'aquest període i, tot i que no hem estudiat quin és l'efecte sobre els resultats obtinguts, creiem convenient remarcar aquest aspecte.

## 9. WEBGRAFIA

- Bernardo, Á. (2013, May 22). El Top 50 de Spotify: Sus canciones más escuchadas y compartidas. Retrieved from <https://hipertextual.com/2013/05/spotify-mostrara-su-top-50>
- Create Elegant Data Visualisations Using the Grammar of Graphics. (n.d.). Retrieved from <https://ggplot2.tidyverse.org/>
- Música. (2020, December 08). Retrieved from <https://ca.wikipedia.org/wiki/Música>
- QQ-plots: Quantile-Quantile plots - R Base Graphs. (n.d.). Retrieved from <http://www.sthda.com/english/wiki/qq-plots-quantile-quantile-plots-r-base-graphs>
- R plot() Function (Add Titles, Labels, Change Colors and Overlaying Pots). (2018, October 08). Retrieved from <https://www.datamentor.io/r-programming/plot-function/>
- Remove Axis Values of Plot in Base R (3 Examples): Delete Axes. (2020, September 30). Retrieved from <https://statisticsglobe.com/remove-axis-values-of-plot-in-r>
- Web API Reference. (n.d.). Retrieved from <https://developer.spotify.com/documentation/web-api/reference/>
- Welcome to Spotipy!¶. (n.d.). Retrieved from <https://spotipy.readthedocs.io/en/2.16.1/>

## 10. ANNEX

### Codi escrit en Python per extreure les dades del Top 50

```
try:
    import spotipy
    import pandas as pd
except:
    print("Install spotipy and pandas libraries:\n$ pip3 install spotipy pandas")
    exit()
from spotipy.oauth2 import SpotifyClientCredentials

client_credentials_manager = SpotifyClientCredentials('id', 'id')
sp =
spotipy.Spotify(client_credentials_manager=client_credentials_manager)

playlist_id = input("Insert playlist URI: ")
playlist_id = playlist_id[17:len(playlist_id)]

content = sp.playlist_items(playlist_id)
playlist_data = []
for i in content['items']:
    track_id = i['track']['id']
    features = sp.audio_features(track_id)
    danceability = features[0]['danceability']
    valence = features[0]['valence']
    energy = features[0]['energy']
    name = i['track']['name']
    popularity = sp.track(track_id)['popularity']
    playlist_data.append([name, popularity, danceability, energy, valence])

df = pd.DataFrame(playlist_data, columns=['name', 'popularity',
'danceability', 'energy', 'valence'])
df.to_csv('data.csv', sep=';', index=False)
```

## Script en R per fer els càlculs i graficar

```
#Introduce data from the csv file.
SpainData<-read.csv(file = "spain.csv", header = TRUE, sep = ";")

#Introduce each row in its respective vector.
energyS<-c(SpainData$energy);
danceabilityS<-c(SpainData$danceability);
valenceS<-c(SpainData$valence);

#Calculate the mean, median and standart deviation from each sample.
meanEnergyS = mean(energyS);
meanDanceabilityS = mean(danceabilityS);
meanValenceS = mean(valenceS);

medianEnergyS = median(energyS);
medianDanceabilityS = median(danceabilityS);
medianValenceS = median(valenceS);

sdEnergyS = sd(energyS);
sdDanceabilityS = sd(danceabilityS);
sdValenceS = sd(valenceS)

#Introduce data from the csv file.
FinlandData<-read.csv(file = "finland.csv", header = TRUE, sep = ";")

#Introduce each row in its respective vector.
energyF<-c(FinlandData$energy);
danceabilityF<-c(FinlandData$danceability);
valenceF<-c(FinlandData$valence);

#Calculate the mean, median and standard deviation from each sample.
meanEnergyF = mean(energyF);
meanDanceabilityF = mean(danceabilityF);
meanValenceF = mean(valenceF);

medianEnergyF = median(energyF);
medianDanceabilityF = median(danceabilityF);
medianValenceF = median(valenceF);

sdEnergyF = sd(energyF);
sdDanceabilityF = sd(danceabilityF);
sdValenceF = sd(valenceF)

#Compute all the Normal plots.
x <- seq(0, 1.3, by = .01);
y <- dnorm(x, meanEnergyF, sdEnergyF);
plot(x,y, type = "l", lwd = 2, axes = TRUE, yaxt = "n", xlab = "", ylab =
"", main = "Energy Normal Distribution")
```

```

x <- seq(0, 1.3, by = .01);
y <- dnorm(x, meanValenceF, sdValenceF);
plot(x,y, type = "l", lwd = 2, axes = TRUE, yaxt = "n", xlab = "", ylab =
"", main = "Valence Normal Distribution")

x <- seq(0, 1.3, by = .01);
y <- dnorm(x, meanDanceabilityF, sdDanceabilityF);
plot(x,y, type = "l", lwd = 2, axes = TRUE, yaxt = "n", xlab = "", ylab =
"", main = "Danceability Normal Distribution")

#Compute all the Boxplots.

boxplot(ylim=c(0.0,1.0),valenceS,valenceF, main = "Valence Boxplot", ylab =
"Valence Values", names=c("Spain","Finland"),col=c("#ff8936","#6f97f2"));
boxplot(ylim=c(0.0,1.0),energyS,energyF, main = "Energy Boxplot", ylab =
"Energy Values", names=c("Spain","Finland"),col=c("#ff8936","#6f97f2"));
boxplot(ylim=c(0.0,1.0),danceabilityS,danceabilityF, main = "Danceability
Boxplot", ylab = "Danceability Values",
names=c("Spain","Finland"),col=c("#ff8936","#6f97f2"));

ddf <- structure(list(energyF, valenceF, danceabilityF), .Names =
c("Energy", "Valence","Danceability"), row.names = c(NA, -5L), class =
"data.frame")
boxplot(ddf[1:3],main = "Energy, Valence and Danceability Boxplots", ylab =
"Stablished Values from 0.0 to 1.0")

#Code for plot all the normals distribution

vectorEnergyS<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:50)
{
  if (energyS[i] >= 0 && energyS[i] < 0.1) vectorEnergyS[1] =
vectorEnergyS[1] +1;
  if (energyS[i] >= 0.1 && energyS[i] < 0.2) vectorEnergyS[2] =
vectorEnergyS[2] +1;
  if (energyS[i] >= 0.2 && energyS[i] < 0.3) vectorEnergyS[3] =
vectorEnergyS[3] +1;
  if (energyS[i] >= 0.3 && energyS[i] < 0.4) vectorEnergyS[4] =
vectorEnergyS[4] +1;
  if (energyS[i] >= 0.4 && energyS[i] < 0.5) vectorEnergyS[5] =
vectorEnergyS[5] +1;
  if (energyS[i] >= 0.5 && energyS[i] < 0.6) vectorEnergyS[6] =
vectorEnergyS[6] +1;
  if (energyS[i] >= 0.6 && energyS[i] < 0.7) vectorEnergyS[7] =
vectorEnergyS[7] +1;

```



```

    if (energyS[i] >= 0.7 && energyS[i] < 0.8) vectorEnergyS[8] =
vectorEnergyS[8] +1;
    if (energyS[i] >= 0.8 && energyS[i] < 0.9) vectorEnergyS[9] =
vectorEnergyS[9] +1;
    if (energyS[i] >= 0.9 && energyS[i] < 1.0) vectorEnergyS[10] =
vectorEnergyS[10] +1;
}

vectorEnergyF<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:50)
{
    if (energyF[i] >= 0 && energyF[i] < 0.1) vectorEnergyF[1] =
vectorEnergyF[1] +1;
    if (energyF[i] >= 0.1 && energyF[i] < 0.2) vectorEnergyF[2] =
vectorEnergyF[2] +1;
    if (energyF[i] >= 0.2 && energyF[i] < 0.3) vectorEnergyF[3] =
vectorEnergyF[3] +1;
    if (energyF[i] >= 0.3 && energyF[i] < 0.4) vectorEnergyF[4] =
vectorEnergyF[4] +1;
    if (energyF[i] >= 0.4 && energyF[i] < 0.5) vectorEnergyF[5] =
vectorEnergyF[5] +1;
    if (energyF[i] >= 0.5 && energyF[i] < 0.6) vectorEnergyF[6] =
vectorEnergyF[6] +1;
    if (energyF[i] >= 0.6 && energyF[i] < 0.7) vectorEnergyF[7] =
vectorEnergyF[7] +1;
    if (energyF[i] >= 0.7 && energyF[i] < 0.8) vectorEnergyF[8] =
vectorEnergyF[8] +1;
    if (energyF[i] >= 0.8 && energyF[i] < 0.9) vectorEnergyF[9] =
vectorEnergyF[9] +1;
    if (energyF[i] >= 0.9 && energyF[i] < 1.0) vectorEnergyF[10] =
vectorEnergyF[10] +1;
}

vectorDanceabilityS<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:50)
{
    if (danceabilityS[i] >= 0 && danceabilityS[i] < 0.1)
vectorDanceabilityS[1] = vectorDanceabilityS[1] +1;
    if (danceabilityS[i] >= 0.1 && danceabilityS[i] < 0.2)
vectorDanceabilityS[2] = vectorDanceabilityS[2] +1;
    if (danceabilityS[i] >= 0.2 && danceabilityS[i] < 0.3)
vectorDanceabilityS[3] = vectorDanceabilityS[3] +1;
    if (danceabilityS[i] >= 0.3 && danceabilityS[i] < 0.4)
vectorDanceabilityS[4] = vectorDanceabilityS[4] +1;
    if (danceabilityS[i] >= 0.4 && danceabilityS[i] < 0.5)
vectorDanceabilityS[5] = vectorDanceabilityS[5] +1;
    if (danceabilityS[i] >= 0.5 && danceabilityS[i] < 0.6)

```

```

vectorDanceabilityS[6] = vectorDanceabilityS[6] +1;
  if (danceabilityS[i] >= 0.6 && danceabilityS[i] < 0.7)
vectorDanceabilityS[7] = vectorDanceabilityS[7] +1;
  if (danceabilityS[i] >= 0.7 && danceabilityS[i] < 0.8)
vectorDanceabilityS[8] = vectorDanceabilityS[8] +1;
  if (danceabilityS[i] >= 0.8 && danceabilityS[i] < 0.9)
vectorDanceabilityS[9] = vectorDanceabilityS[9] +1;
  if (danceabilityS[i] >= 0.9 && danceabilityS[i] < 1.0)
vectorDanceabilityS[10] = vectorDanceabilityS[10] +1;
}

vectorDanceabilityF<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:50)
{
  if (danceabilityF[i] >= 0 && danceabilityF[i] < 0.1) vectorDanceabilityF[1]
= vectorDanceabilityF[1] +1;
  if (danceabilityF[i] >= 0.1 && danceabilityF[i] < 0.2)
vectorDanceabilityF[2] = vectorDanceabilityF[2] +1;
  if (danceabilityF[i] >= 0.2 && danceabilityF[i] < 0.3)
vectorDanceabilityF[3] = vectorDanceabilityF[3] +1;
  if (danceabilityF[i] >= 0.3 && danceabilityF[i] < 0.4)
vectorDanceabilityF[4] = vectorDanceabilityF[4] +1;
  if (danceabilityF[i] >= 0.4 && danceabilityF[i] < 0.5)
vectorDanceabilityF[5] = vectorDanceabilityF[5] +1;
  if (danceabilityF[i] >= 0.5 && danceabilityF[i] < 0.6)
vectorDanceabilityF[6] = vectorDanceabilityF[6] +1;
  if (danceabilityF[i] >= 0.6 && danceabilityF[i] < 0.7)
vectorDanceabilityF[7] = vectorDanceabilityF[7] +1;
  if (danceabilityF[i] >= 0.7 && danceabilityF[i] < 0.8)
vectorDanceabilityF[8] = vectorDanceabilityF[8] +1;
  if (danceabilityF[i] >= 0.8 && danceabilityF[i] < 0.9)
vectorDanceabilityF[9] = vectorDanceabilityF[9] +1;
  if (danceabilityF[i] >= 0.9 && danceabilityF[i] < 1.0)
vectorDanceabilityF[10] = vectorDanceabilityF[10] +1;
}

vectorValenceS<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:50)
{
  if (valenceS[i] >= 0 && valenceS[i] < 0.1) vectorValenceS[1] =
vectorValenceS[1] +1;
  if (valenceS[i] >= 0.1 && valenceS[i] < 0.2) vectorValenceS[2] =
vectorValenceS[2] +1;
  if (valenceS[i] >= 0.2 && valenceS[i] < 0.3) vectorValenceS[3] =
vectorValenceS[3] +1;
  if (valenceS[i] >= 0.3 && valenceS[i] < 0.4) vectorValenceS[4] =
vectorValenceS[4] +1;

```

```

    if (valenceS[i] >= 0.4 && valenceS[i] < 0.5) vectorValenceS[5] =
vectorValenceS[5] +1;
    if (valenceS[i] >= 0.5 && valenceS[i] < 0.6) vectorValenceS[6] =
vectorValenceS[6] +1;
    if (valenceS[i] >= 0.6 && valenceS[i] < 0.7) vectorValenceS[7] =
vectorValenceS[7] +1;
    if (valenceS[i] >= 0.7 && valenceS[i] < 0.8) vectorValenceS[8] =
vectorValenceS[8] +1;
    if (valenceS[i] >= 0.8 && valenceS[i] < 0.9) vectorValenceS[9] =
vectorValenceS[9] +1;
    if (valenceS[i] >= 0.9 && valenceS[i] < 1.0) vectorValenceS[10] =
vectorValenceS[10] +1;
}

vectorValenceF<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:50)
{
    if (valenceF[i] >= 0 && valenceF[i] < 0.1) vectorValenceF[1] =
vectorValenceF[1] +1;
    if (valenceF[i] >= 0.1 && valenceF[i] < 0.2) vectorValenceF[2] =
vectorValenceF[2] +1;
    if (valenceF[i] >= 0.2 && valenceF[i] < 0.3) vectorValenceF[3] =
vectorValenceF[3] +1;
    if (valenceF[i] >= 0.3 && valenceF[i] < 0.4) vectorValenceF[4] =
vectorValenceF[4] +1;
    if (valenceF[i] >= 0.4 && valenceF[i] < 0.5) vectorValenceF[5] =
vectorValenceF[5] +1;
    if (valenceF[i] >= 0.5 && valenceF[i] < 0.6) vectorValenceF[6] =
vectorValenceF[6] +1;
    if (valenceF[i] >= 0.6 && valenceF[i] < 0.7) vectorValenceF[7] =
vectorValenceF[7] +1;
    if (valenceF[i] >= 0.7 && valenceF[i] < 0.8) vectorValenceF[8] =
vectorValenceF[8] +1;
    if (valenceF[i] >= 0.8 && valenceF[i] < 0.9) vectorValenceF[9] =
vectorValenceF[9] +1;
    if (valenceF[i] >= 0.9 && valenceF[i] < 1.0) vectorValenceF[10] =
vectorValenceF[10] +1;
}

#Do all t-test
t.test(spainData$valence, finlandData$valence, alternative = "greater", mu =
0, paired=FALSE, var.equal=TRUE, conf.level = 0.99)

t.test(spainData$energy, finlandData$energy, alternative = "greater", mu =
0, paired=FALSE, var.equal=TRUE, conf.level = 0.99)

t.test(spainData$danceability, finlandData$danceability, alternative =
"greater", mu = 0, paired=FALSE, var.equal=TRUE, conf.level = 0.99)

```

```

#Draw all the barplots
barra_x <-c("0.1","0.2","0.3","0.4","0.5","0.6","0.7","0.8","0.9","1.0");

barplot(xlab = "Energy Values",ylab =
"Frequency",ylim=c(0,20),vectorEnergyS, main = "Spain Top 50 Energy
Distribution",names.arg=barra_x, col = "#ff8936");
barplot(xlab = "Energy Values",ylab =
"Frequency",ylim=c(0,20),vectorEnergyF,main = "Finland Top 50 Energy
Distribution",names.arg=barra_x,col="#6f97f2");
barplot(xlab = "Danceability Values",ylab =
"Frequency",ylim=c(0,20),vectorDanceabilityS,main = "Spain Top 50
Danceability Distribution",names.arg=barra_x, col = "#ff8936");
barplot(xlab = "Danceability Values",ylab =
"Frequency",ylim=c(0,20),vectorDanceabilityF,main = "Finland Top 50
Danceability Distribution",names.arg=barra_x,col="#6f97f2");
barplot(xlab = "Valence Values",ylab =
"Frequency",ylim=c(0,20),vectorValenceS,main = "Spain Top 50 Valence
Distribution",names.arg=barra_x, col = "#ff8936");
barplot(xlab = "Valence Values",ylab =
"Frequency",ylim=c(0,20),vectorValenceF,main = "Finland Top 50 Valence
Distribution",names.arg=barra_x,col="#6f97f2");

#Draw all the QQplots
qqnorm(energyS, pch = 1, frame = FALSE,main = "Q-Q plot from Spain's Energy
Values",ylim = c(0,1.0))
qqline(energyS, col = "#ff8936", lwd = 2)
qqnorm(energyF, pch = 1, frame = FALSE,main = "Q-Q plot from Finland's
Energy Values",ylim = c(0,1.0))
qqline(energyF, col = "#6f97f2", lwd = 2)

qqnorm(valenceS, pch = 1, frame = FALSE,main = "Q-Q plot from Spain's
Valence Values",ylim = c(0,1.0))
qqline(valenceS, col = "#ff8936", lwd = 2)
qqnorm(valenceF, pch = 1, frame = FALSE,main = "Q-Q plot from Finland's
Valence Values",ylim = c(0,1.0))
qqline(valenceF, col = "#6f97f2", lwd = 2)

qqnorm(danceabilityS, pch = 1, frame = FALSE,main = "Q-Q plot from Spain's
Danceability Values",ylim = c(0,1.0))
qqline(danceabilityS, col = "#ff8936", lwd = 2)
qqnorm(danceabilityF, pch = 1, frame = FALSE,main = "Q-Q plot from Finland's
Danceability Values",ylim = c(0,1.0))
qqline(danceabilityF, col = "#6f97f2", lwd = 2 )

```

XAVIER COLL RIBAS  
ENRIC CONDAL ASENSIO  
ROC SALVADOR ANDREAZINI