A Visualization Study of:

Module Name: Fundamentals of Information Visualisation (G53FIV)
Student Name: Evi Nikolaidou, Student ID: 4336685

Introduction

The Eurovision Song Contest is an international song competition, among countries mainly in European Broadcast Union (EBU). Initially the competition was called "The Eurovision Grand Prix", and it was inspired by the famous Italian Festival, San Remo.

"At the beginning of the television age in the early 1950's, the European television landscape offered the image of a political divided and technical fragmented continent" (Esc-history.com, 2019). At the same time, EBU was looking for a way to reunite the European Countries after the disruption of World War II. In 1955, Marcel Bezençon, came up with the idea of the Eurovision Song Contests. Back then, it was a technologically challenging project, to join that many countries together in a wide area international network.

The first Eurovision Song Contest was held in Lugano, Switzerland in 1953 and with only 7 participants. "With a live orchestra, the norm in the early years, and simple sing-along songs on every radio station, the Contest grew into a true pan-European tradition" (Eurovision.tv, 2019). The contest grown through the years, nowadays more than 40 countries are participating every year. Also, non- European countries like Israel and in recent years Australia are also participating.

Eurovision Song Contest was recognized by the Guinness Book of World Record as the Longest Running Annual TV Music Competition in 2015. During the last 62 years more than 1500 songs was produced from more than 50 countries, including countries that do not exist anymore (e.g. Yugoslavia).

Every year each participating country submit one original song and performs it on a live television and radio. Then each country is televoting the favourite song and the winning country would host the contest the next year. Throughout the years many changes have happened on the voting system. In earlier years, only juries could vote and there was no televoting from the public but now its 50/50 between juries and telephone votes. Moreover, due the growing number of the participation demand, semi-finals were introduced in 2007. Beside changes in voting system, changes happened and for the presentations of the votes. Ever since 1994, each country announces their results by having their spokesperson standing in front of a famous landmark of their country. Currently the voting and presentation system works like this: The viewers can vote by telephone or by the official mobile app. Each national spokesperson will announce the results of their professional jury in the usual way by giving points 1-8, 10 and 12 to their favourites. Then the televoting points from each country will be combined and announce by the host starting by the country that received the less points from the public and ending with the country that received the most points. Therefore, the winner will be announced at the very end of the show.

In this report, a dataset from the official Eurovision website that contains information mainly about the winners of the Eurovision Song Contest, from 1956 to 2018 was collected. Based on this dataset, five questions were formed and answered by visualizing the dataset in an exploration fashion using R language.

Data Description

In the following table you can see the data description of the Eurovision Dataset:

Name	Nominal, Ordinal & Quantitative	Dimensions & Measures
Year	Q- Internal	Dimensions
Win Country	N-Nominal	Dimensions
Song	N-Nominal	Dimensions
Performer	N-Nominal	Dimensions
Start Number	O- Ordered	Measures
Gender	N-Nominal	Dimensions
Song Language	N-Nominal	Dimensions
Points	Q-Ratio	Measures
Runner Up	N-Nominal	Dimensions
Last Place	N-Nominal	Dimensions
No. Participants	Q-Ratio	Measures

Initial Questions

The first initial questions that will be answered are:

1) Is there a secret recipe to win Eurovision?
2) Who is the most loved Eurovision Winner?
3) Does the Performance order number affect the Winner?
4) Which countries almost won most times?
5) How did the participants change through the years?

Setup

In this project the following libraries were used:

```
1. library(dplyr)
2. library(ggplot2)
3. library(scales)
```

dplyr: This library was used for data manipulation.
ggplot2: This library was used for plotting graphs.
scales: This library was used for scaling the x, y axis of the graphs.

Question 1: Is there a secret recipe to win Eurovision?

The first question is about some aspects that may be critical in order to win the Eurovision Song Contest. Starting of, we need a representation of the countries that are most successful in the competition. Then we can check whether the gender and song language are affecting the winner. Is there a combination of those two aspects, song language and gender that was more successful?

To choose the appropriate representation for answering this question and the following once, was to look at figure 1:

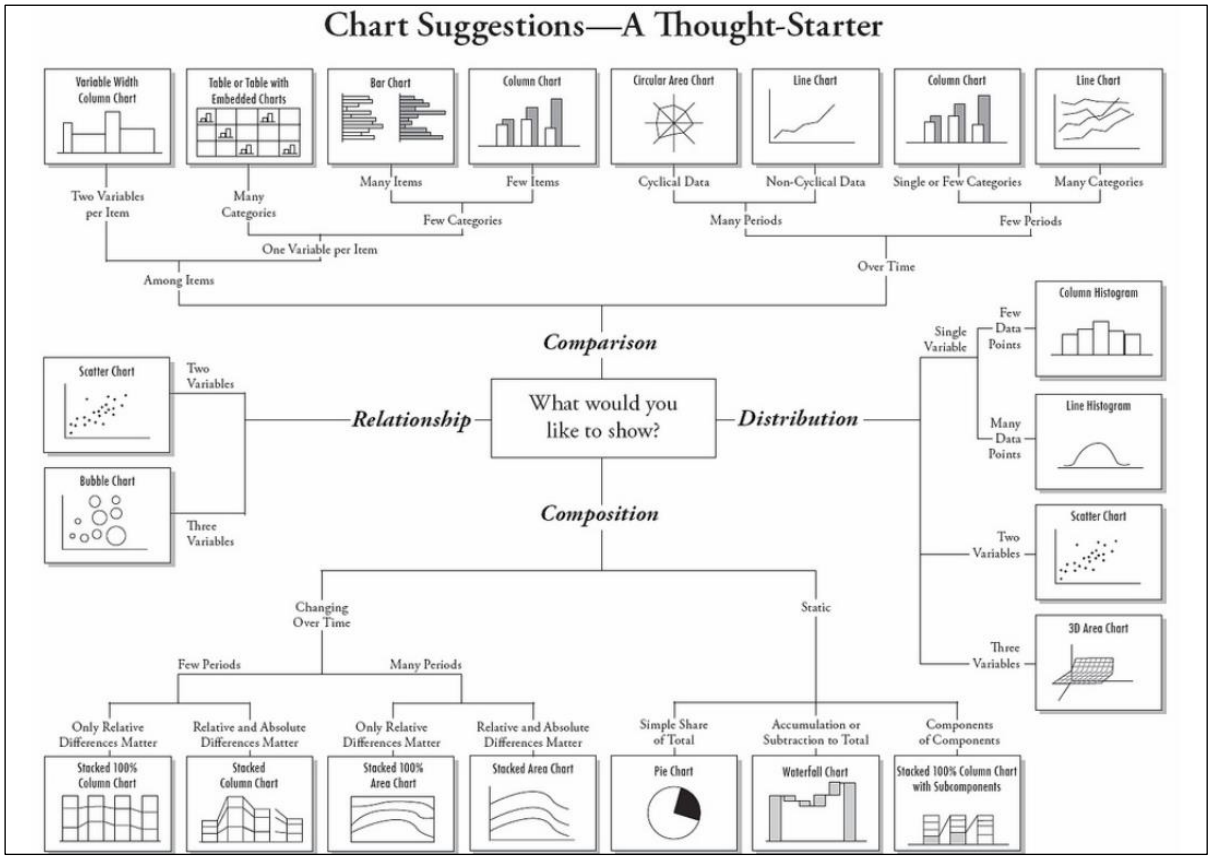
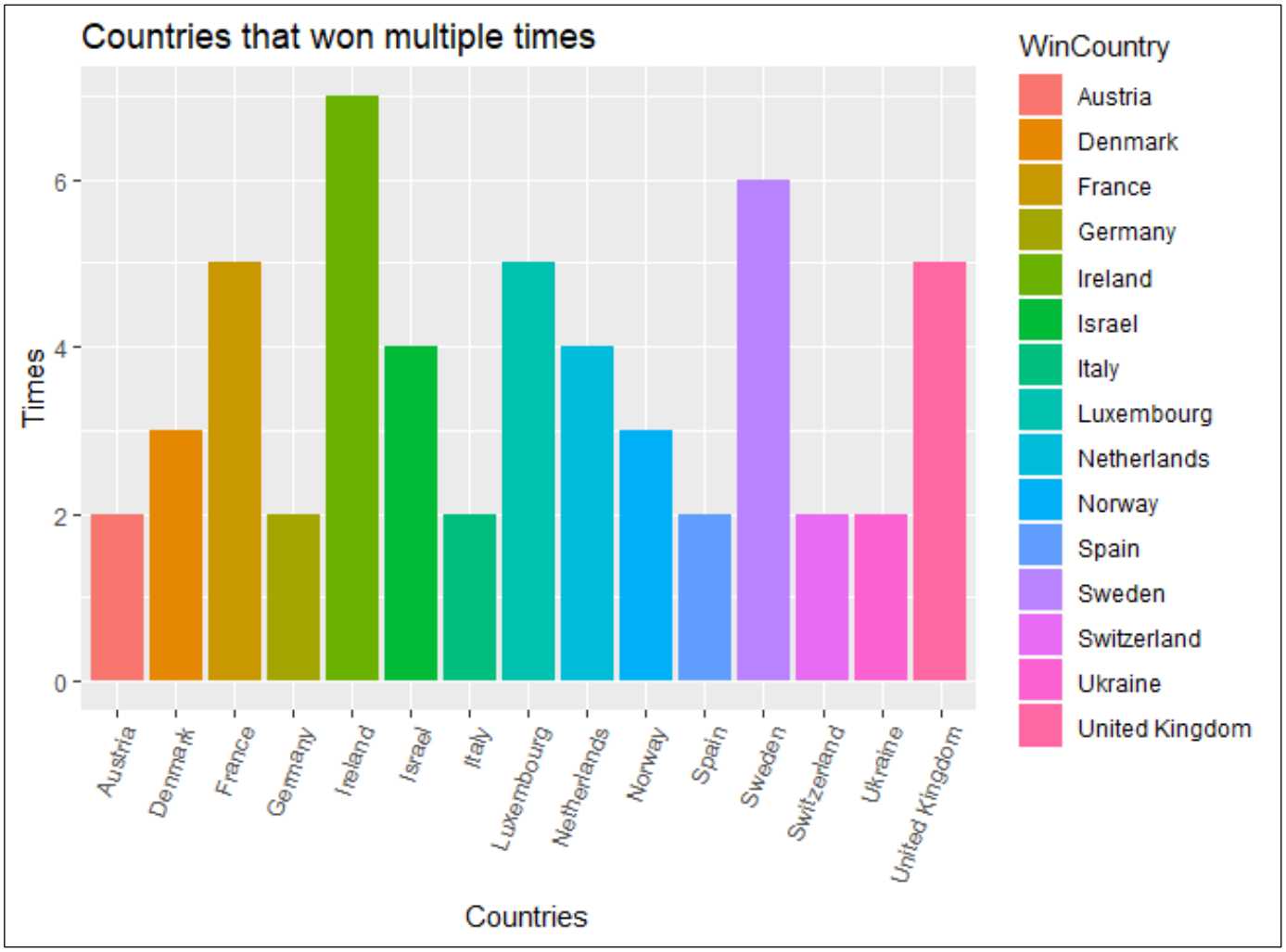


Figure 1: Selecting the appropriate Chart

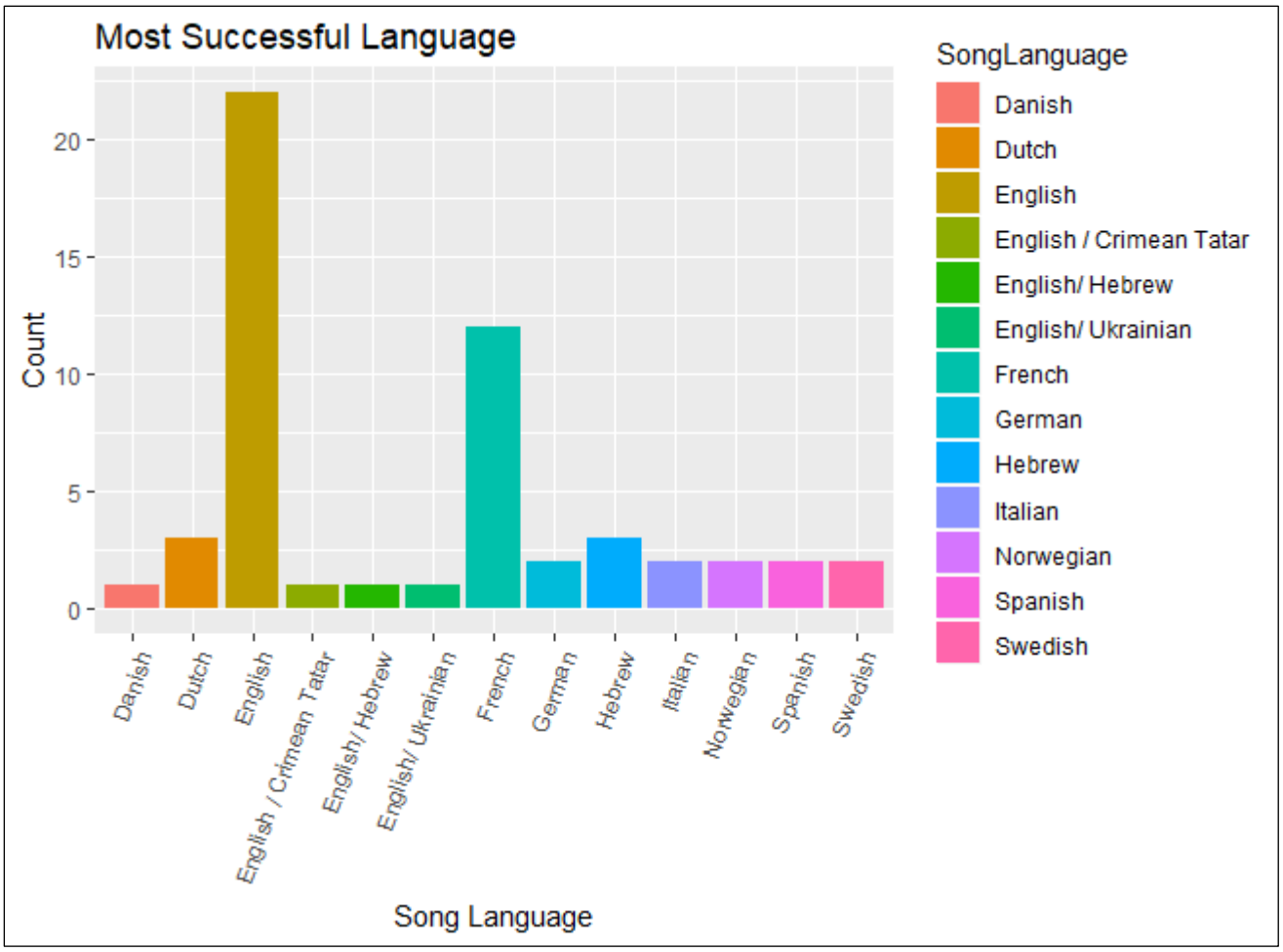
For the question 1, the bar chart was the most appropriate one, because it has only one variable, the winning country but many categories. Thus, let's take a look at the most successful countries in the Competition. To do that we first need to filter the countries that have won more than once (check code, line 3):

```
1.  eurovision %>%
2.    group_by(WinCountry)%>%
3.    filter(n()>1) %>%
4.    ggplot(aes(x=WinCountry, fill=WinCountry)) +
5.    geom_bar() +
6.    xlab("Countries") + ylab("Times") + ggtitle("Countries that won multiple times")+
7.    theme(axis.text.x = element_text(angle = 70, hjust = 1))
```



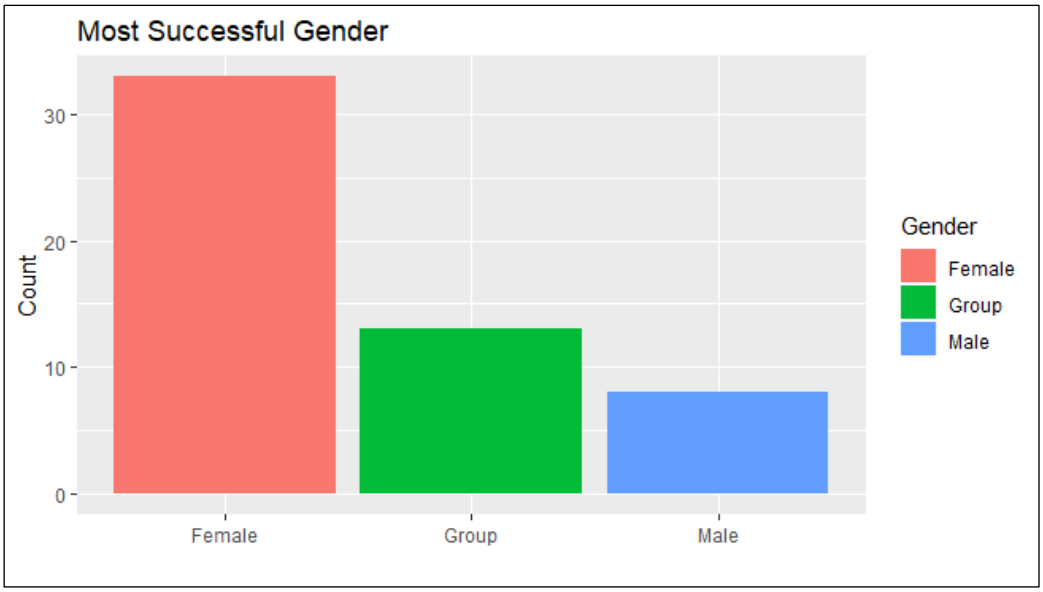
As you can see the most successful country is Ireland and the following is Sweden. Now that we have the most successful countries in the competition, we can assess another question: Which language is most successful between those countries that won most times?

```
eurovision %>%
  group_by(WinCountry)%>%
  filter(n()>1) %>%
  ggplot(aes(x=SongLanguage, fill=SongLanguage)) +
  geom_bar() +
  xlab("Song Language") + ylab("Count") + ggtitle("Most Successful Language")+
  theme(axis.text.x = element_text(angle = 70, hjust = 1))
```



There is a clear instigation that the songs in English are the most successful in the competition. Second more successful are songs in French. Now let's see if the gender can affect the winner. Keeping in mind that some winners were in a mixed group, all groups are considering a different category. Hence, we will have, female, male and Groups. Moreover, for the purpose of this gender-based visualization, Dana International (1998) and Conchita Wurst (2014) are both categorized female in this report.

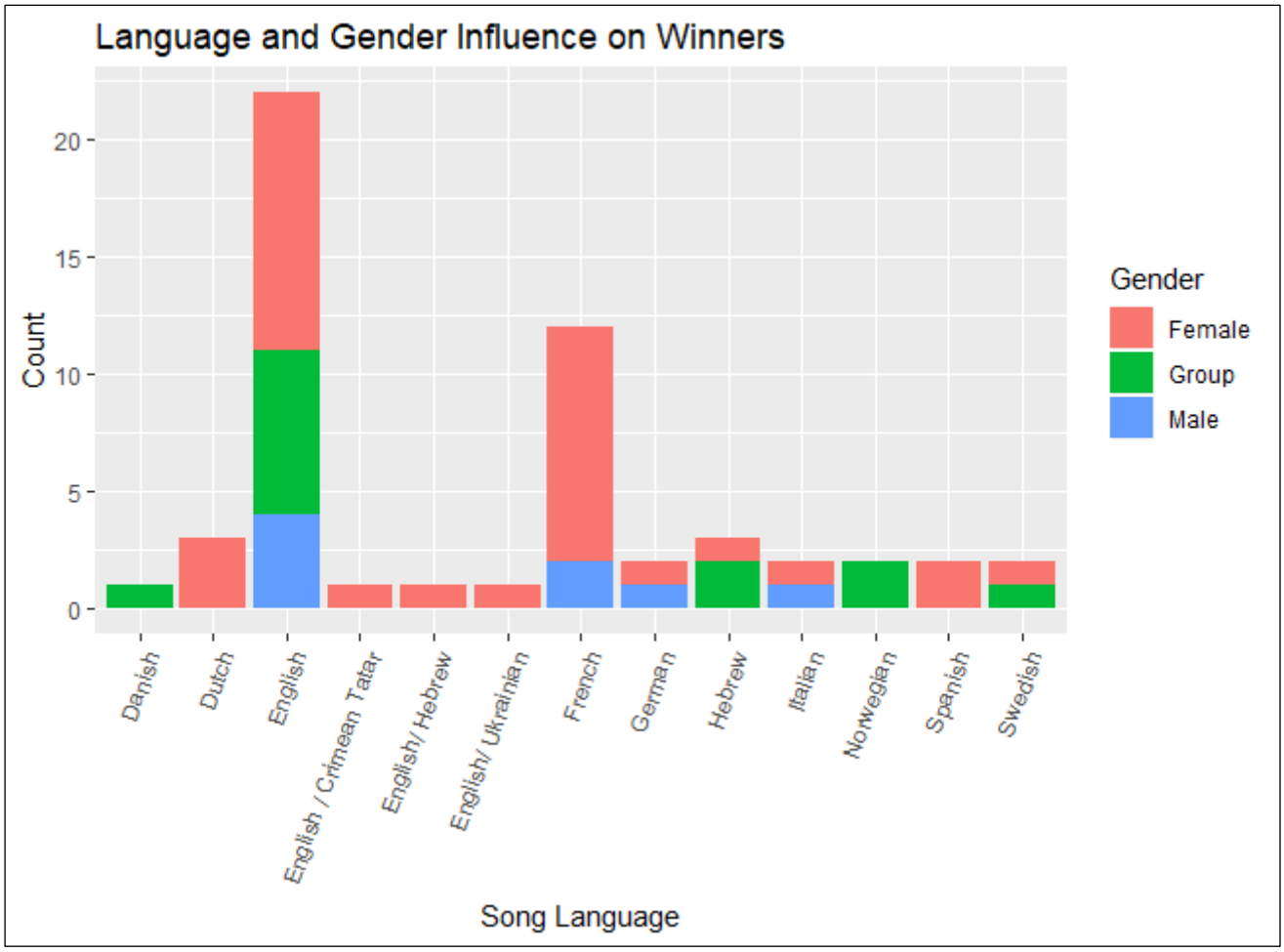
```
eurovision %>%
  group_by(WinCountry)%>%
  filter(n()>1) %>%
  ggplot(aes(x=Gender, fill= Gender)) +
  geom_bar() +
  xlab("") + ylab("Count") + ggtitle("Most Successful Gender")
```



The above chart proves that they are mostly female winners in the Eurovision Song Contest.

Lastly, by combining the two charts above we can have a clear view if the Song Language and the gender of the performer has any significant advantage in winning the competition. To do that, we can use the following code:

```
eurovision %>%
  group_by(WinCountry)%>%
  filter(n()>1) %>%
  ggplot(aes(x=SongLanguage, fill=Gender)) +
  geom_bar() +
  xlab("Song Language") + ylab("Count") + ggtitle("Language and Gender Influence on Winners")+
  theme(axis.text.x = element_text(angle = 70, hjust = 1))
```



From the above chart we can figure that an English Song with a female performer is more likely to win according to the previous winners. Also, there was mostly female artists that won the competition with a French song, only groups that won with a Danish and Norwegian songs. There also no male dominated song language.

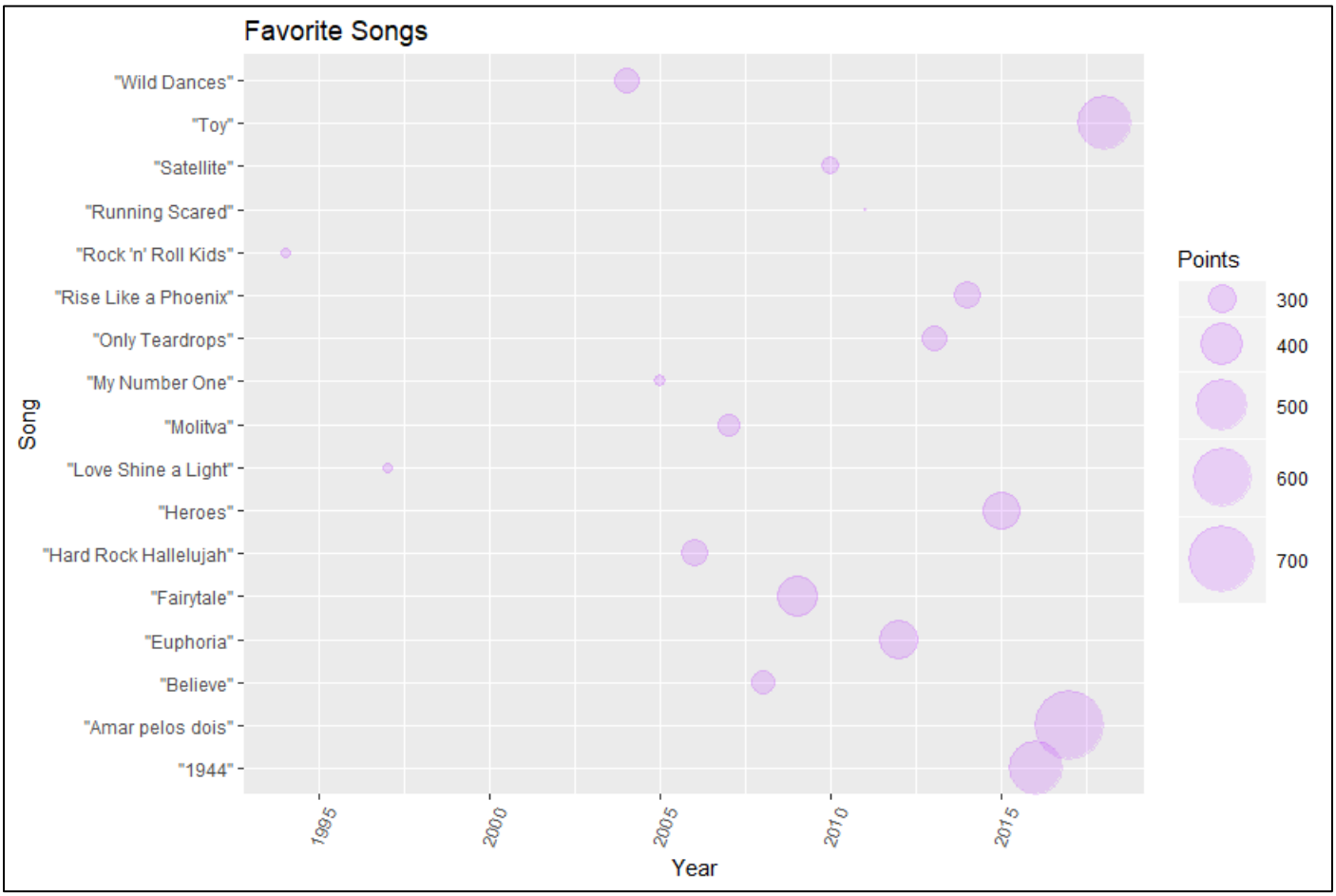
Question 2: Who is the most loved Eurovision Winner?

The second question it covers the most successful winner song of the competition. To figure the most loved Eurovision Songs we need to check the points of each winner.

For this question, we have three variables that need to take into consideration before moving forward. We have the Song, which is nominal, the year that was introduced which is Interval, and the points which is q-radio. The reason we need the year is due to the difference in the voting system throughout the years that affected the number of points. For this question the most appropriate visualization strategy is the bubble plot. It is like Scatter plot (that we will use in a later question), but this allows as to represent 2 numerical variables and a third variable to scale the size of the points.

So first, let's find the most loved winner throughout all years of Eurovision. However, we need to filter out the year 1956 which was the fist year of the competition and no points were revealed. This could result to too many data on the graph and considering we are looking for only the most loved ones, the ones with most points another filter was created. Hence, we filter out the winners with less than 200 points.

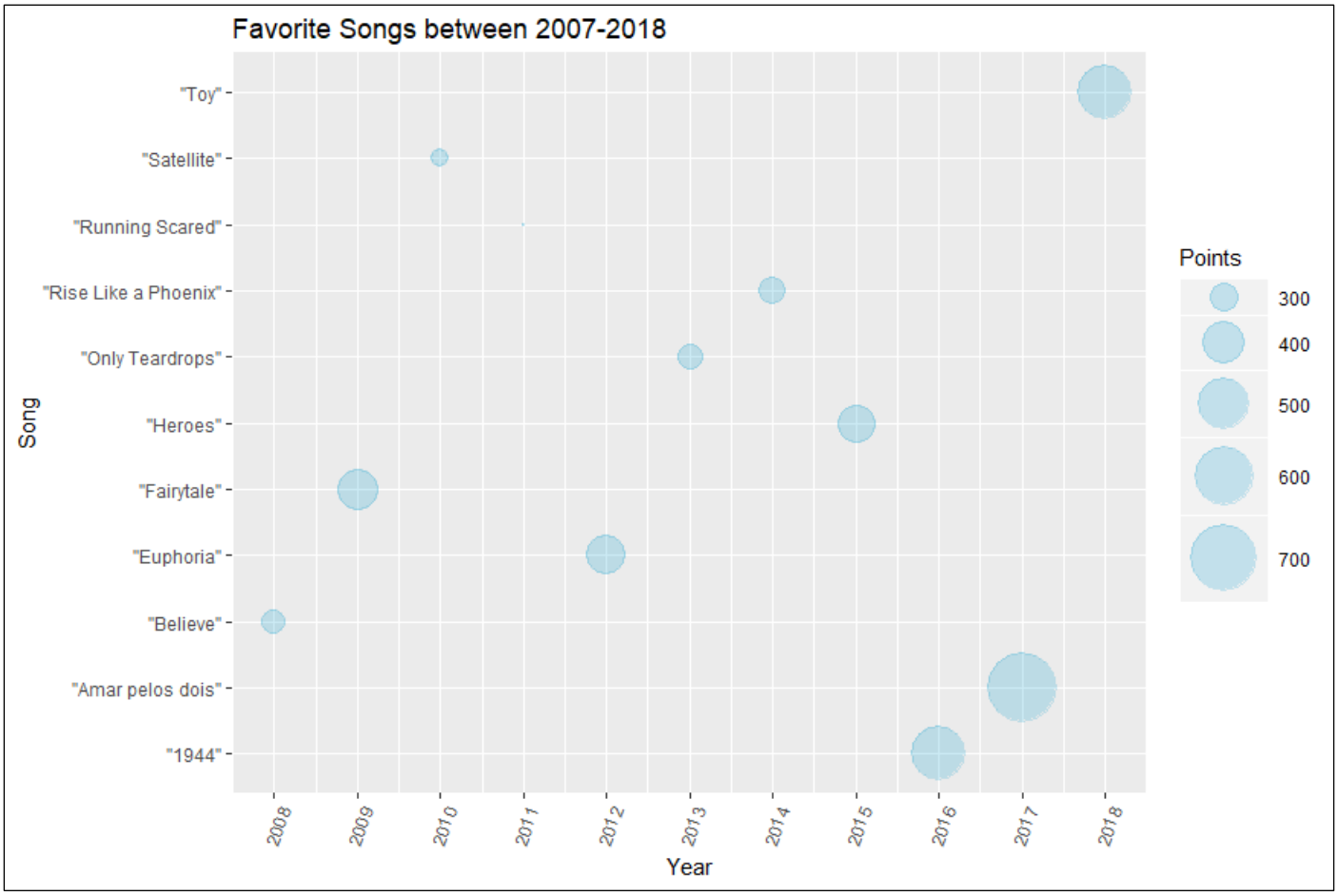
```
eurovision %>%
  filter(Year != 1956 & Points >= 200) %>%
  ggplot(aes(x= Year, y= Song, size= Points)) +
  geom_point (alpha= 0.2 ,color = "darkorchid1") +
  scale_size_continuous(range = c(0.5, 16))+
  xlab("Year") + ylab("Song") + ggtitle("Favorite Songs")+
  theme(axis.text.x = element_text(angle = 70, hjust = 1)) +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 5))
```



From the above visualization we can see that the most successful winners are "Amar pelos dois" followed by "1944" and then "Toy".

Now let's analyse this a bit more. We need to manipulate the data in to show as the winners of the year 2007 until 2018. The reason why these particular years it's because in 2007 the semi-finals started.

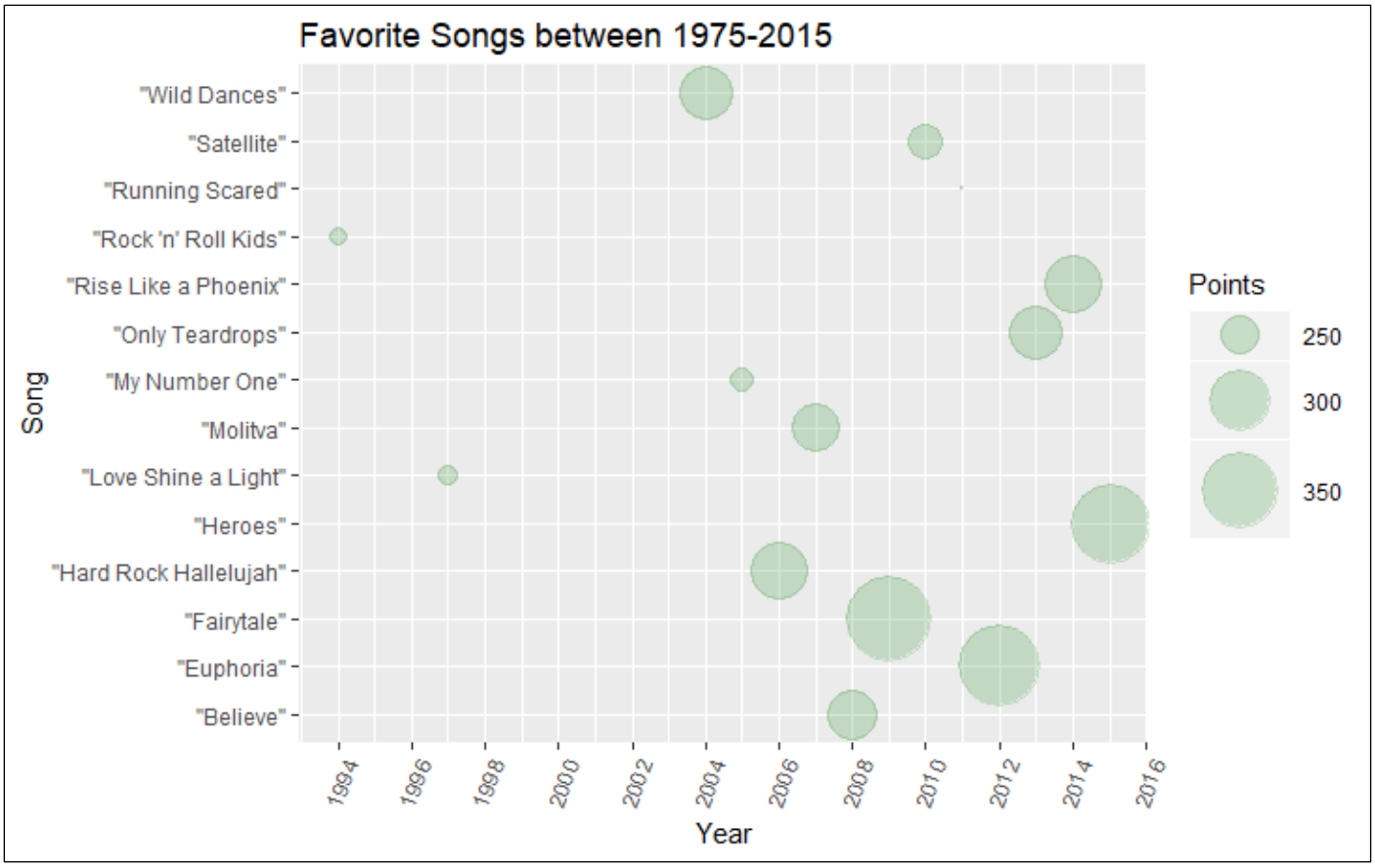
```
eurovision %>%
  filter(Year > 2007 & Points >= 200) %>%
  ggplot(aes(x= Year, y= Song, size= Points)) +
  geom_point (alpha= 0.2 ,color = "deepskyblue3") +
  scale_size_continuous(range = c(0.5, 16))+
  xlab("Year") + ylab("Song") + ggtitle("Favorite Songs between 2007-2018")+
  theme(axis.text.x = element_text(angle = 70, hjust = 1)) +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
```



This did not change any of the results, again the most successful winners are "Amar pelos dois" followed by "1944" and then "Toy".

As we can see the three most successful winners are in the years from 2016 until today. The reason why is because the voting system change that year. Therefore, we should present only the winners that belong to the same voting system to be fairer. The years between 1975 – 2015 have the same voting system.

```
eurovision %>%
  filter(Year >= 1975 & Year <= 2015  & Points >= 200) %>%
  ggplot(aes(x= Year, y= Song, size= Points)) +
  geom_point (alpha= 0.2 ,color = "forestgreen") +
  scale_size_continuous(range = c(0.5, 16))+
  xlab("Year") + ylab("Song") + ggtitle("Favorite Songs between 1975-2015")+
  theme(axis.text.x = element_text(angle = 70, hjust = 1)) +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
```



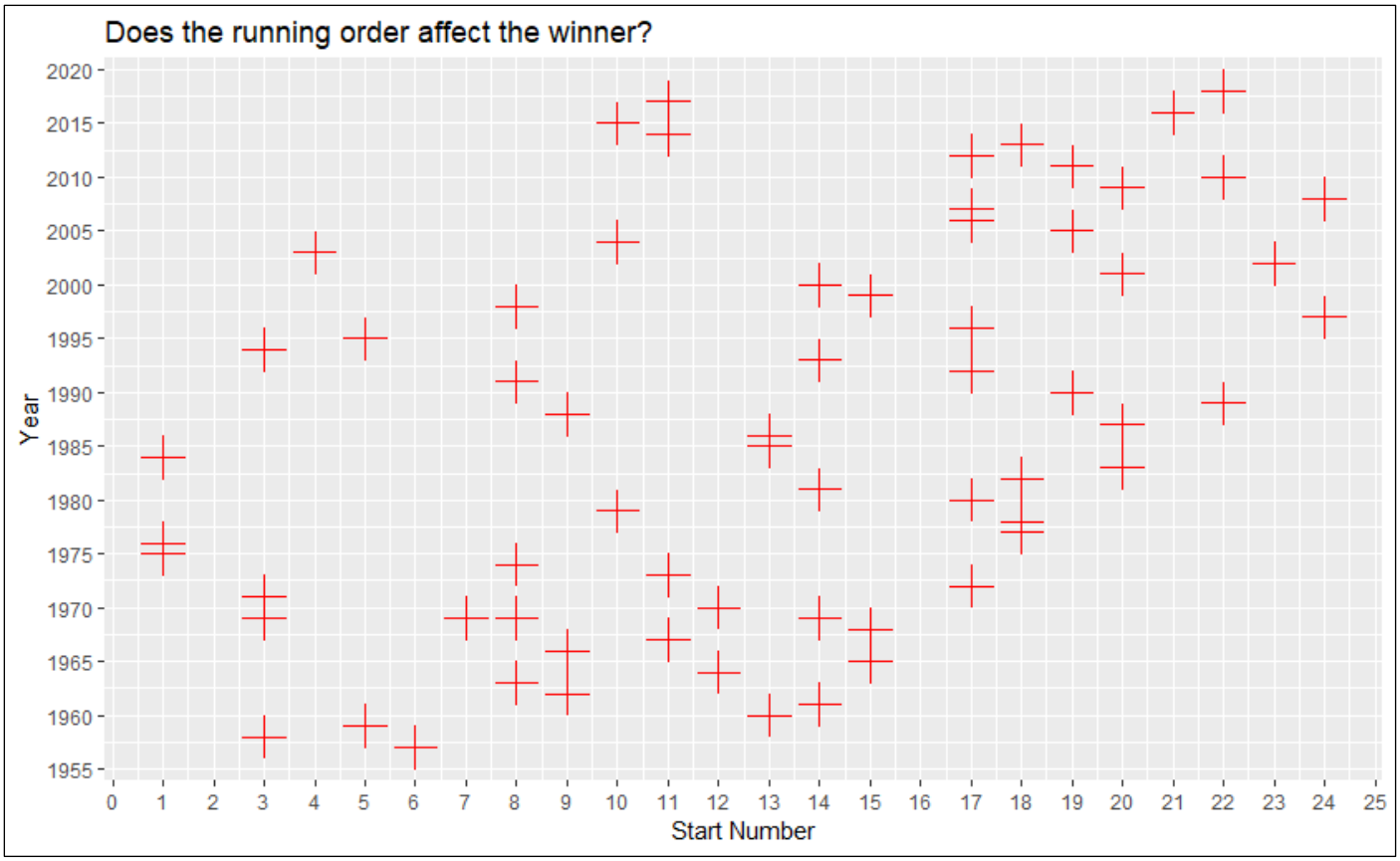
Now the results have dramatically change. The most successful winner is “Fairytale” followed by “Euphoria” and “Heroes”. The difference in the points of Euphoria and Heroes is too small and its not easy to read from the chart which has most points.

Question 3: Does the Performance order number affect the Winner?

There is an alleged “curse” that specify that the if you perform second in the competition you will not win. Hence, this question will explore which start numbers are the luckiest and which are the unluckiest according to the start numbers of the previous winners.

For this question we have two variables, the year and the start number, also known as the performance order number. For this reason, the most appropriate visualization strategy is the Scatter plot. Similar to the bubble plot but only allows display of the value of 2 set of data on 2 dimensions. However, we need to filter out the year 1956 which was the first year of the competition and no results were revealed.

```
1.  eurovision %>%
2.    filter(Year != 1956) %>%
3.    ggplot(aes(x= StartNumber, y= Year)) +
4.    geom_point(cex=6, pch=3, alpha=1.0, color= "Red") +
5.    scale_x_continuous(breaks = scales::pretty_breaks(n = 26)) +
6.    scale_y_continuous(breaks = scales::pretty_breaks(n = 15)) +
7.    xlab("Start Number")+ ylab("Year")+ ggtitle("Does the running order affect the winner?")
```



The above chart "proves" the curse as nobody that performed 2nd has ever won, but also in place 7th, 16th and 25th and up. But this is acceptable as usually no more than 26 country are performing in the final.

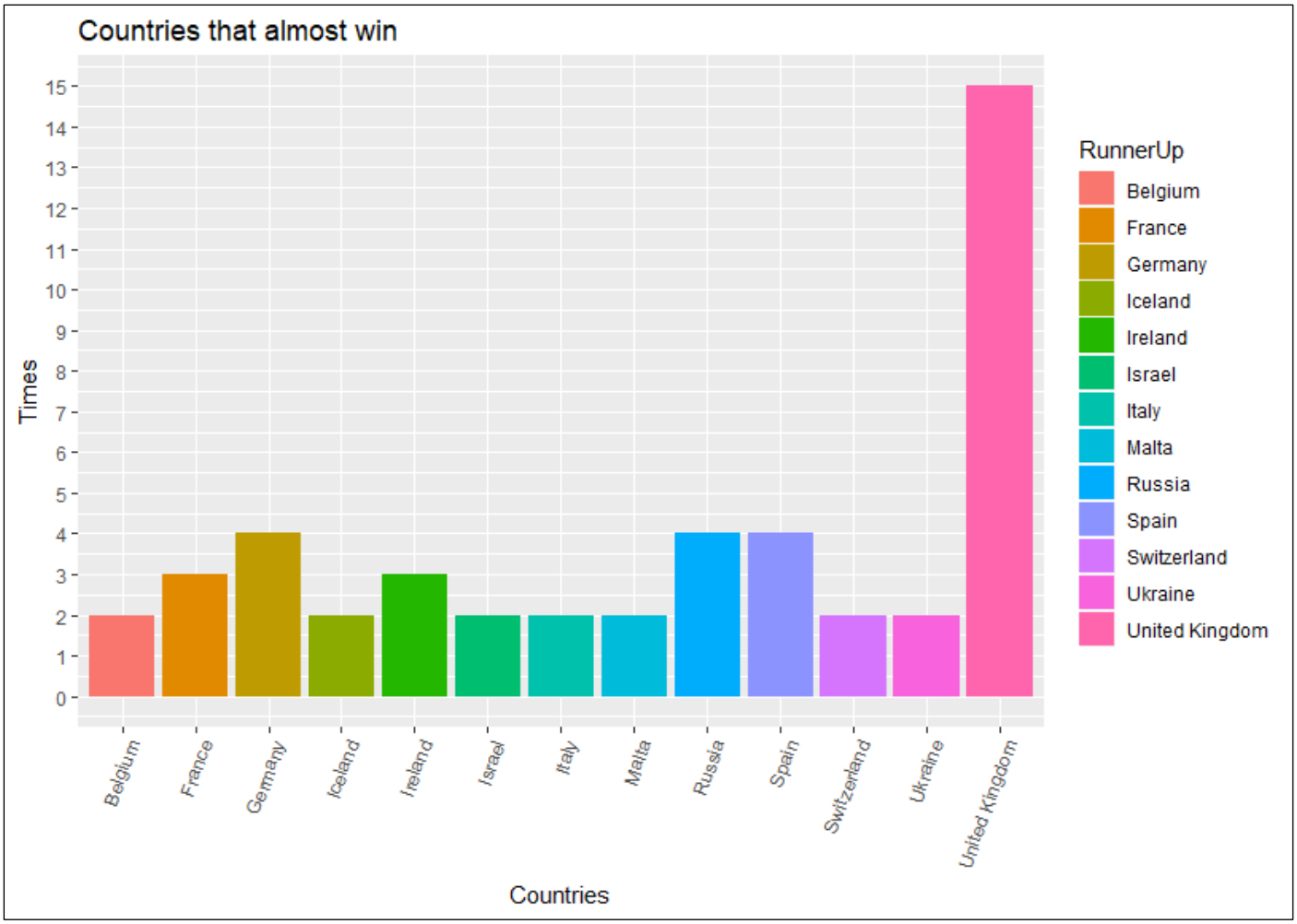
Question 4: Which countries almost won most times?

Every year, there is country that almost won. In this question we analyse the country that came second in the competition most times.

Again, for this question, the bar chart was the most appropriate one, because it has only one variable, the runner up country also known as the country that took the second place in the competition.

However, we need to filter out the year 1956 which was the first year of the competition and no results were revealed and the year 1969 where four countries (Spain, United Kingdom, Netherlands and France) won the competition.

```
eurovision %>%
  group_by(RunnerUp)%>%
  filter(n()>1) %>%
  filter(Year != 1956 & Year != 1969) %>%
  ggplot(aes(x=RunnerUp, fill=RunnerUp)) +
  geom_bar() +
  xlab("Countries") + ylab("Times") + ggtitle("Countries that almost win")+
  theme(axis.text.x = element_text(angle = 70, hjust = 1))+
  scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
```



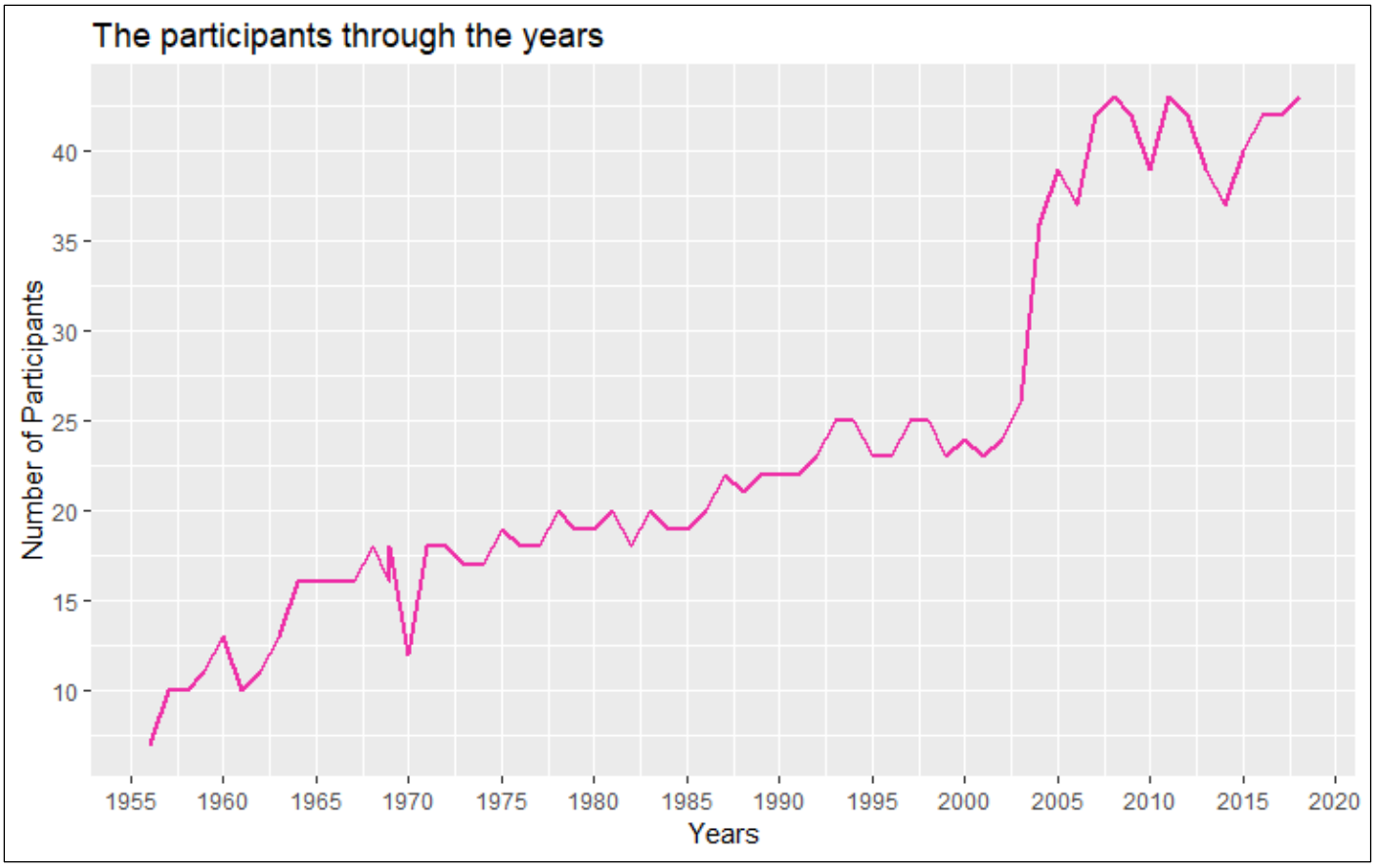
The results show that the United Kingdom was by far the Runner Up most times. It's interesting that if you compared the results of the countries that won most times (question 1) and the countries that almost won (question 4), France, Germany, Ireland, Italy, Spain, Switzerland, Ukraine and United Kingdom are in both lists. Therefore, those are most successful in the competition. Another interesting point is that beside Ireland and Switzerland, the rest are the Big Five. Big Five are the countries that are automatically qualify for the final without going through the semi-finals, because those are the five biggest financial contributors to the Eurovision Song Contest.

Question 5: How did the participants changed through the years?

The first ever competition had only 7 participants and last year the were 43 participants. Thought the years some countries stop existing, split, change names and other stop participating. How did that change through the years? How long did it take for more countries to start participating in the competition?

For this question, the most appropriate visualization was the line plot because it represents the relationship between two variables. In this case we need to explore the relationship between the number participants and the years.

```
ggplot(eurovision, aes(Year,NoParticipants))+
  geom_line( colour = "maroon2", size = 1) +
  xlab("Years") + ylab("Number of Participants")+ ggtitle("The participants through the years")+
  scale_x_continuous(breaks = scales::pretty_breaks(n = 10)) +
  scale_y_continuous(breaks = scales::pretty_breaks(n = 10))
```



The chart above can tell as that although there was a significant increase in the overall number of the participants there was some time were there was a drop. For example, in 1970 there was a clear drop of 10 participants and a big increase during the 2000s.

Reflection of the development Process

Personal Reflection:

I found this assessment challenging but at the same time extremely interesting. It gave me the opportunity to analyse, a dataset of my choice which was interesting to me and expand my knowledge on the fundamentals of the data visualization. Throughout this project I gain valuable information and understanding on some of the visualization strategies that I had to follow in order to complete this project. Choosing the Eurovision Song Contest for a dataset not only encourage me to implement a visualization but also explore other technics that I could include.

Challenges, Limitations and Future Work

One of the biggest challenges that I could do differently is that I should create more functions. Specifically, instead of writing the plots for each question a better more readable code would include functions for each plot and calling it when I need it instead of creating the same plot multiple times.

Another issue that I had to face was for the second question. By importing the Eurovision.csv file from the script, R studio was reading the column “Points” as a character instead of numeric. Therefore, when I try to filter the points the results were wrong. The dataset was an interesting subject, for me that I have some knowledge of, so I realize that the most favourite songs was wrong because some of them were missing. I soon realise that If I import the dataset from the R studio option you can change the points from character to numeric.

Last by using a dataset that included only the winners of Eurovision, limited the possible questions that I could explore. In the future, more datasets for Eurovision including all the participants could create more visualizations with more depth.

References

- Fenn, D., Suleman, O., Efstathiou, J. and Johnson, N. (2006). How does Europe Make Its Mind Up? Connections, cliques, and compatibility between countries in the Eurovision Song Contest. Physica A: Statistical Mechanics and its Applications, 360(2), pp.576-598.
- Esc-history.com. (2019). Eurovision Song Contest : How it all began : ESC-History. [online] Available at: https://www.esc-history.com/ [Accessed 7 Apr. 2019].
- Eurovision.tv. (2019). In a Nutshell - Eurovision Song Contest. [online] Available at: https://eurovision.tv/history/in-a-nutshell [Accessed 7 Apr. 2019].
- Eurovision.tv. (2019). Facts & Figures - Eurovision Song Contest. [online] Available at: https://eurovision.tv/about/facts-and-figures [Accessed 7 Apr. 2019].
- Eurovision.tv. (2019). Biggest change to Eurovision Song Contest voting since 1975. [online] Available at: https://eurovision.tv/story/biggest-change-to-eurovision-song-contest-voting-since-1975 [Accessed 7 Apr. 2019].