Data Visualization Project

Kalmanidis Theodoros
ID:218580

Language Used: R

The dataset

The dataset we are going to use in this project is called metal_bands_2017.csv and it consists of band information from all over the world from 1960 to 2016 .

The objective

The objective of this assignment is to simplify and increase comprehensibility and communication efficiency on existing visualizations for the white metal_bands_2017.csv dataset. Existing visualizations for this dataset we are going to use are located publicly on https://data.world/ and https://www.kaggle.com and have been created either wirh R or Python.

The variables

Band_name: The name of band

Fans: Number of the known fans

Formed: Formation year of the band

Origin:Location where the band was formed

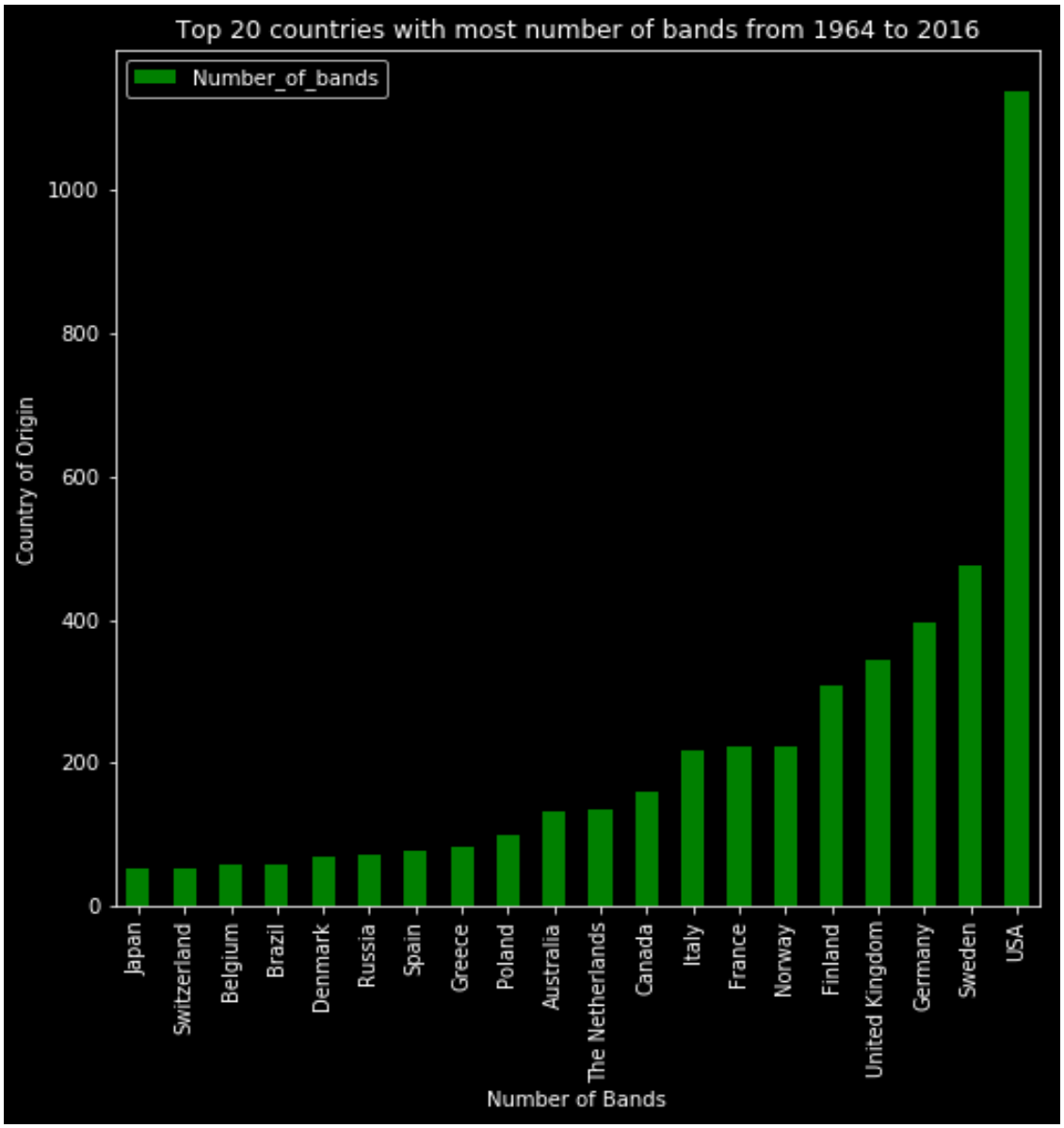
Split: The year the band splitted

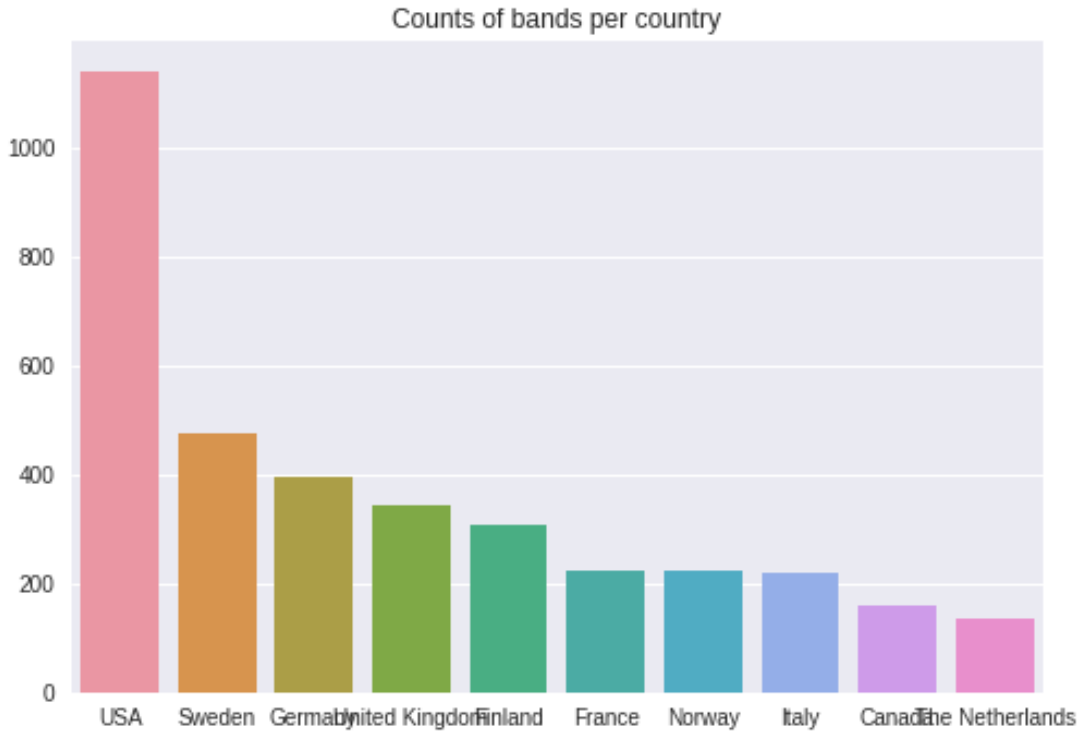
Style: The music style the band was playing

Data Visualizations

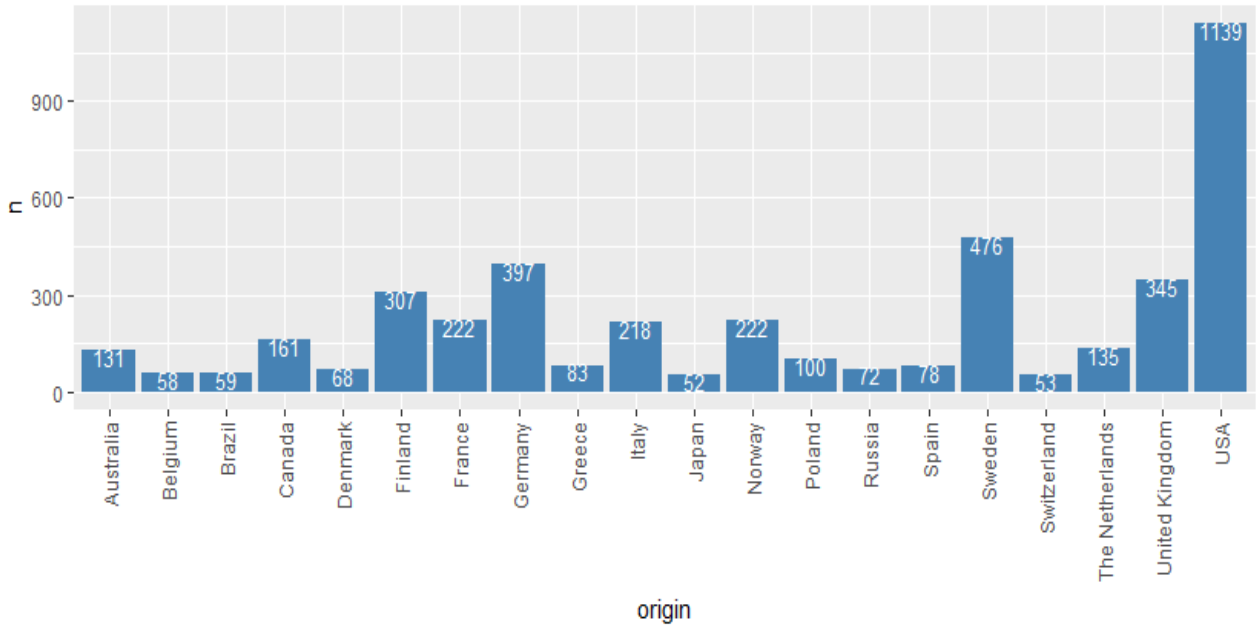
In this part we are going to improve already existing visualizations for this dataset and discuss the cons and the pros after the improvements(using R) of each one

In the first found barplot for this dataset we can see that the colours used are not very pleasing to the eye and also is no so easy to match is bar with its value.

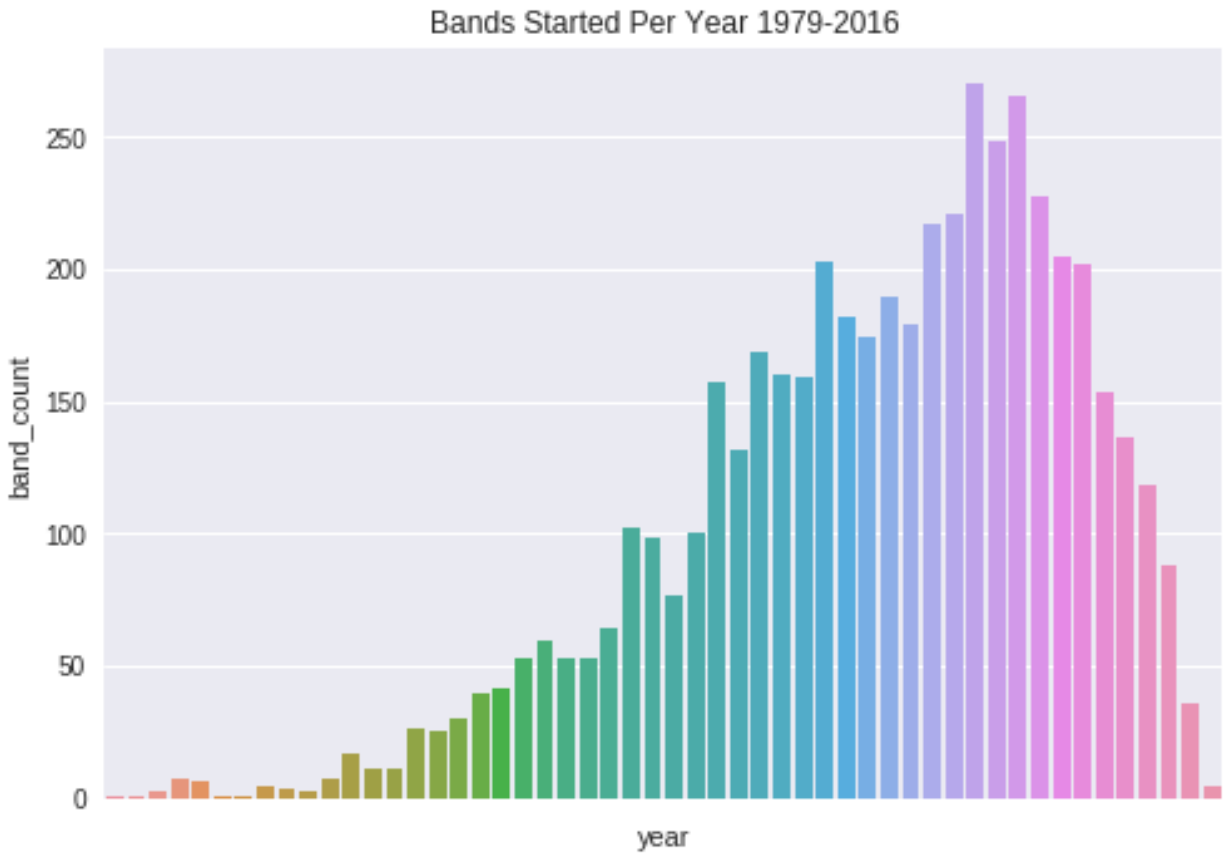




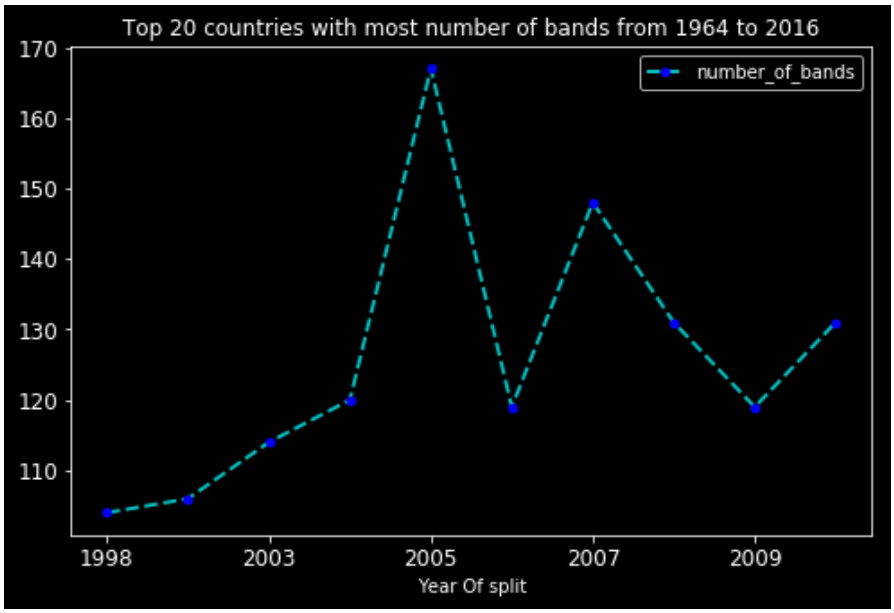
In the second graph above we see that the names are overlapping, although the colours used are better compared to the first graph. So we solve the weaknesses of the previous two using light friendly colours and also plotting the value of each bar inside the bar.



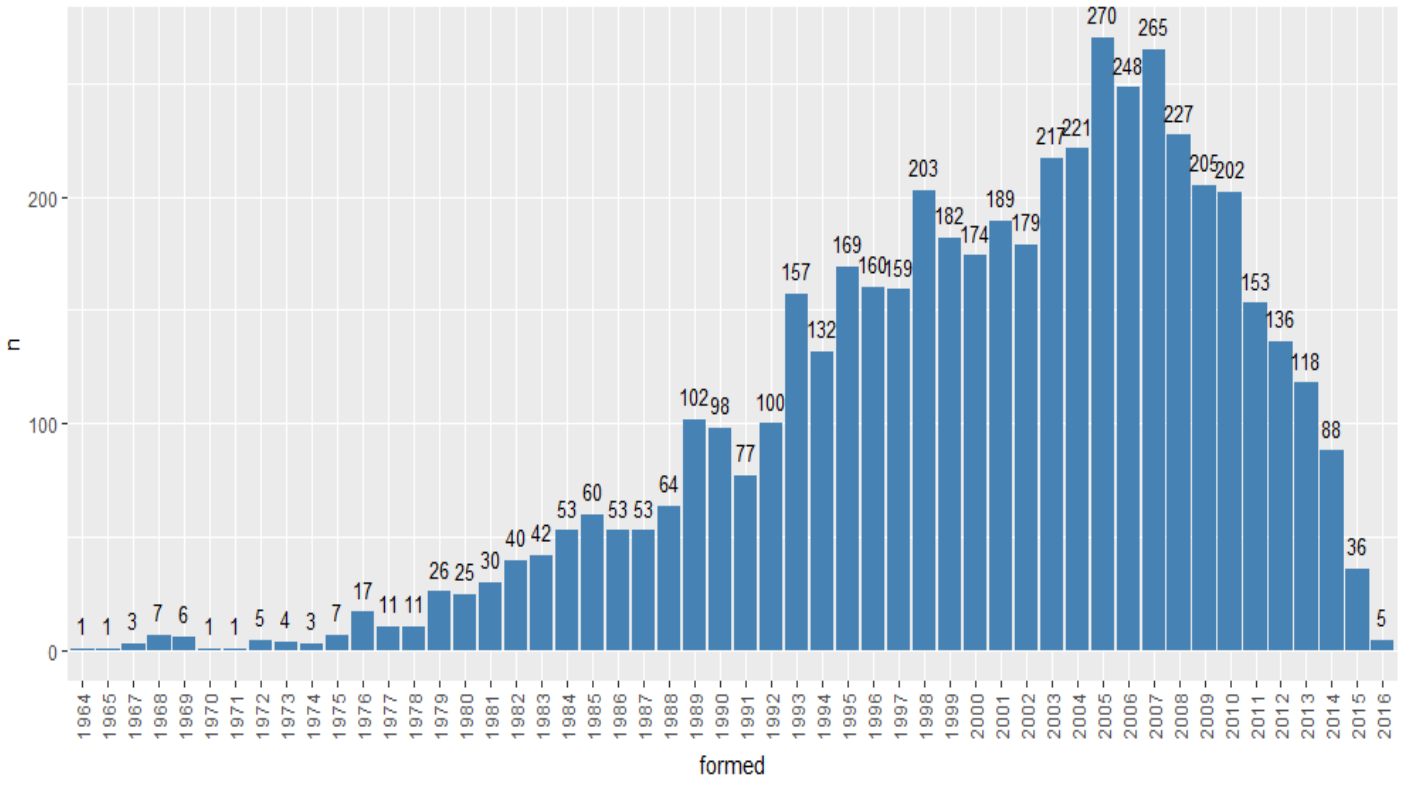
In the third graph we see a band count for each year but we don't see the year listed on x axis or the matching value of each bar with the y axis. It seems the barplot used like a histogram to show the overall tendency which is wrong in this case.



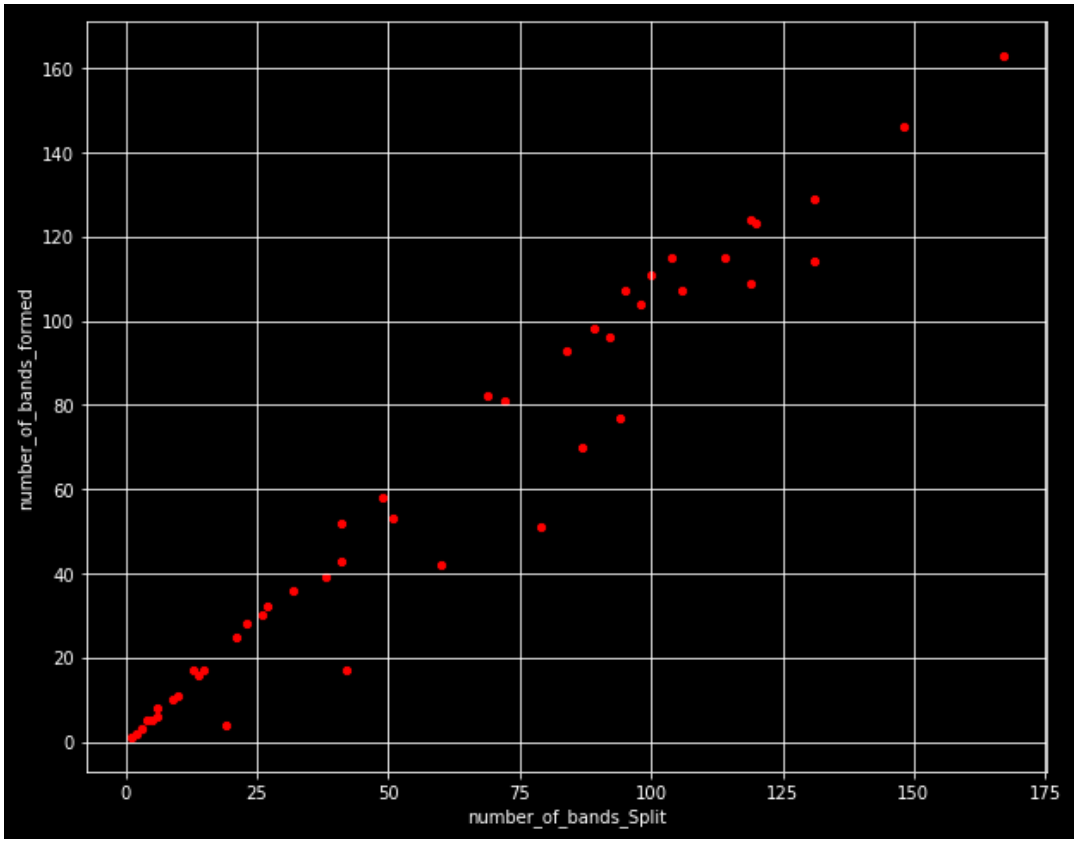
In the fourth graph appeared in the next page which tries to depict the same information, we see again some tendency between years but what if someones wants to know the exact values between the years appeared? This graph hided some truthfulness.

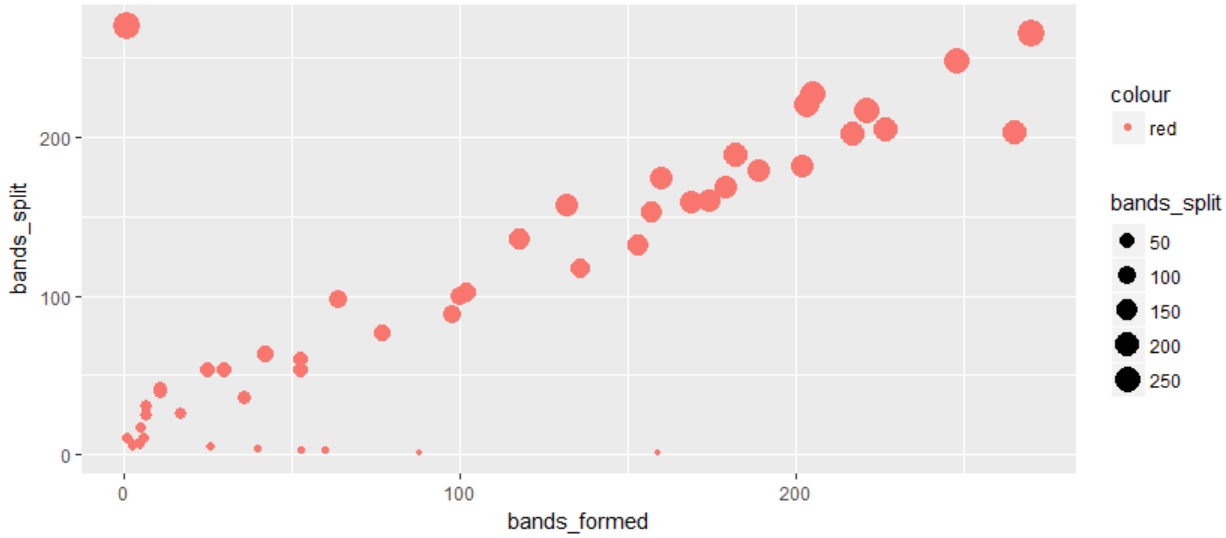


Below we see the improved barplot we construct, using again values over each bar plus all the year listed in a row.



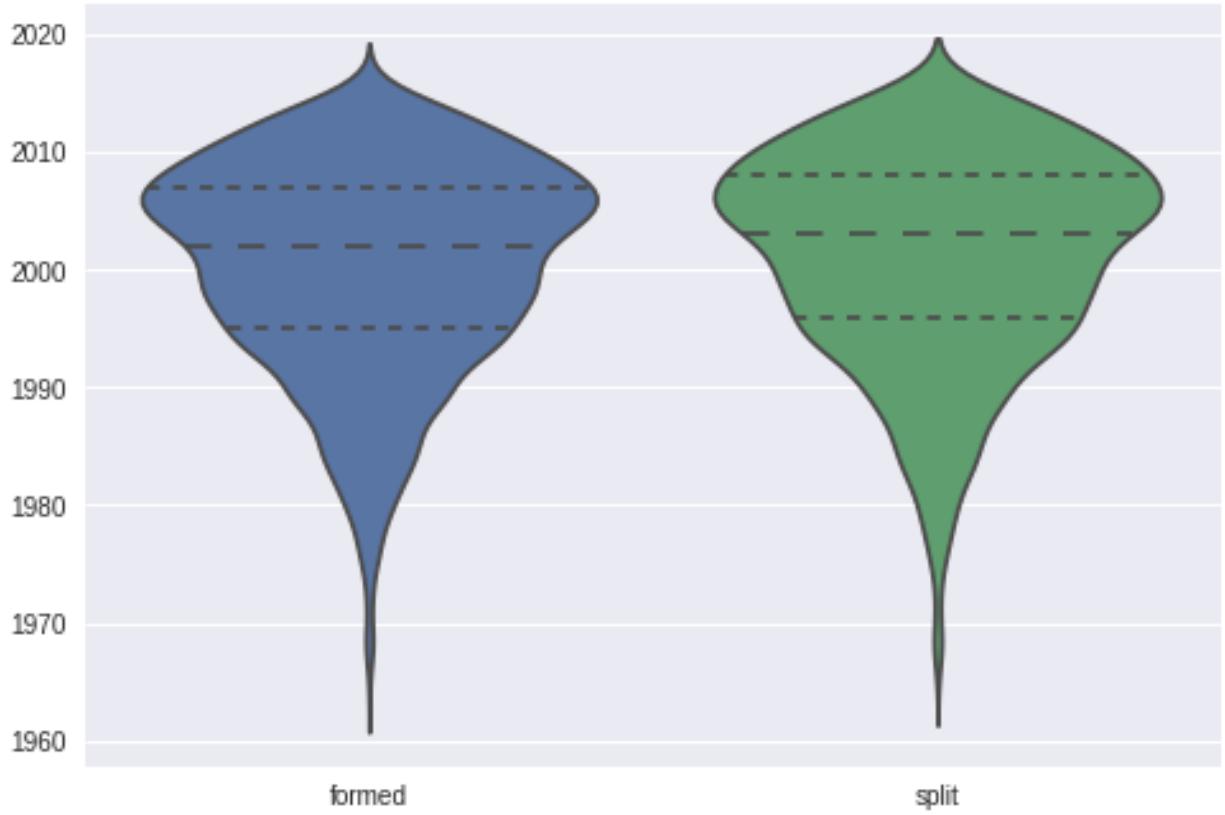
The next graph is a scatterplot of number of bands formed vs band splitted again we see wrong choice of colours the combination of blank font with red point is too aggressive for the human eye



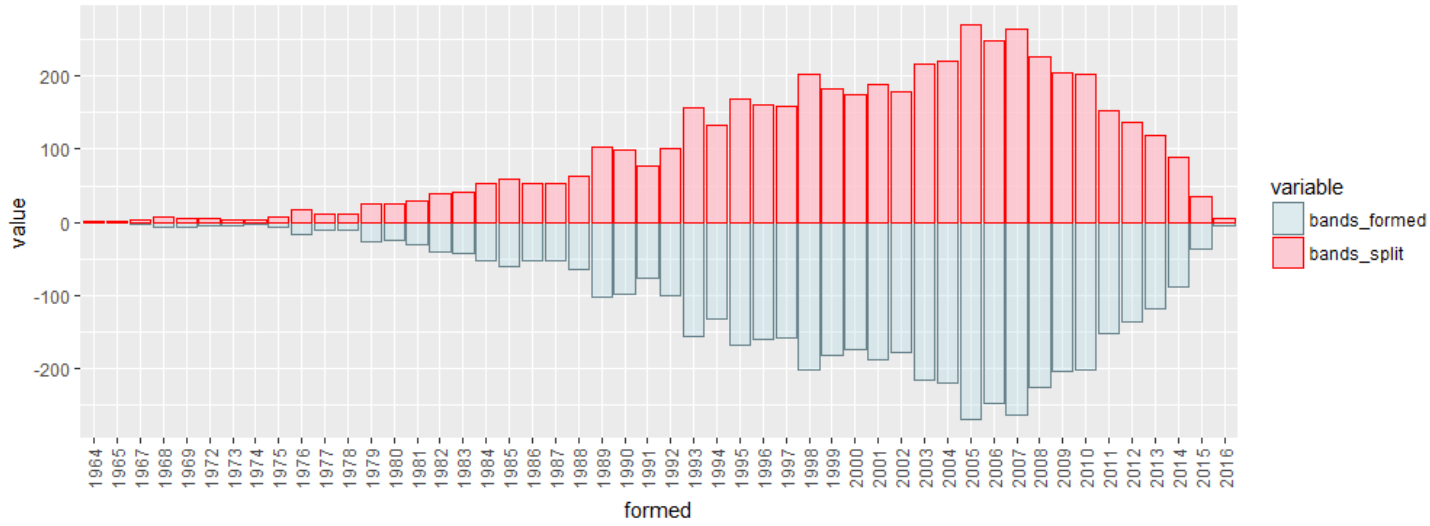


Below we see the improved scatterplot we suggest, not only it uses better colours for the eye view but also different sizes for each point in order to get better understanding of the scatterplot.

In the last found plot for this dataset we see two violin plots one for bands formed and one for bands split. The purpose is to show the constrast of those two for each year but due to the similar size of the violin plots and also the choice of bright colours its not good for most of the viewers.



We know from theory that the best graph for comparison is most of the times barplot. So as we can see on the next page, we choose grouped constrasted barplots with opposite colours in order to help the viewer to notice the contrast. Notice all the years added and not only the decaded like the violin plots, for better precision to the comparison.


value
200
100
0
-100
-200
1964
1965
1967
1968
1969
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
formed
variable
bands_formed
bands_split