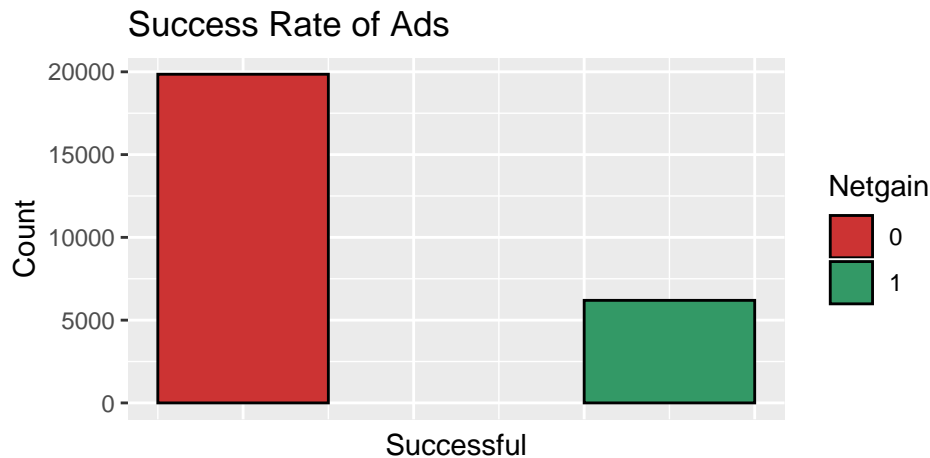


# Advertisement Success - EDA

This project evaluates a data set containing various advertisements to determine which variables are important in creating a successful Ad.

The dataset contains 12 variables in total. This included a unique ID (which will not be used in the model). The response variable is a binary output which determines whether the ad had a net gain or net loss. This was coded as a “1” if the ad had a net gain and “0” if the ad had a net loss.



Around 30% of the Ads were successful

## Getting a glimpse of the data

Now we try to determine the type of our variables and group them into categorical, ordinal and continuous. This is done in order to perform the EDA.

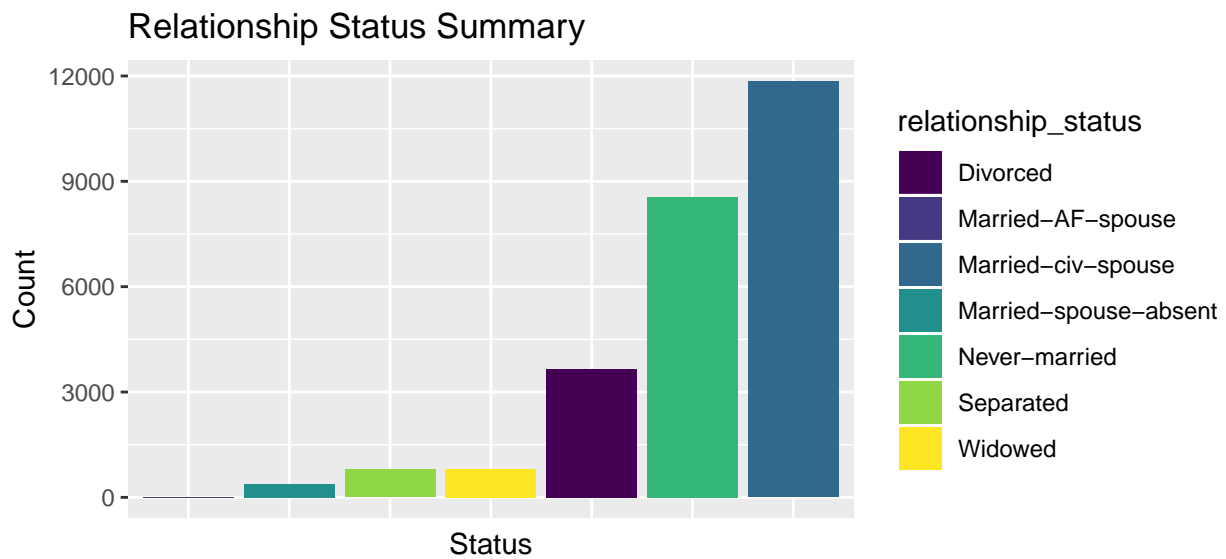
```
## Observations: 26,048
## Variables: 12
## $ id <int> 19717, 31593, 5681, 15491, 23587,...
## $ relationship_status <fct> Married-spouse-absent, Married-ci...
## $ industry <fct> Auto, Pharma, Entertainment, Poli...
## $ genre <fct> Comedy, Comedy, Comedy, Infomerci...
## $ targeted_sex <fct> Male, Male, Female, Female, Male,...
## $ average_runtime.minutes_per_week. <int> 45, 45, 45, 40, 48, 40, 50, 40, 4...
## $ airtime <fct> Primetime, Primetime, Primetime, ...
## $ airlocation <fct> United-States, United-States, Uni...
## $ ratings <dbl> 0.02746467, 0.02746467, 0.0274646...
## $ expensive <fct> High, Low, High, Low, High, Low, ...
## $ money_back_guarantee <fct> No, No, Yes, No, No, Yes, No, Yes...
## $ netgain <dbl> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, ...
```

Here are the following types of variables:

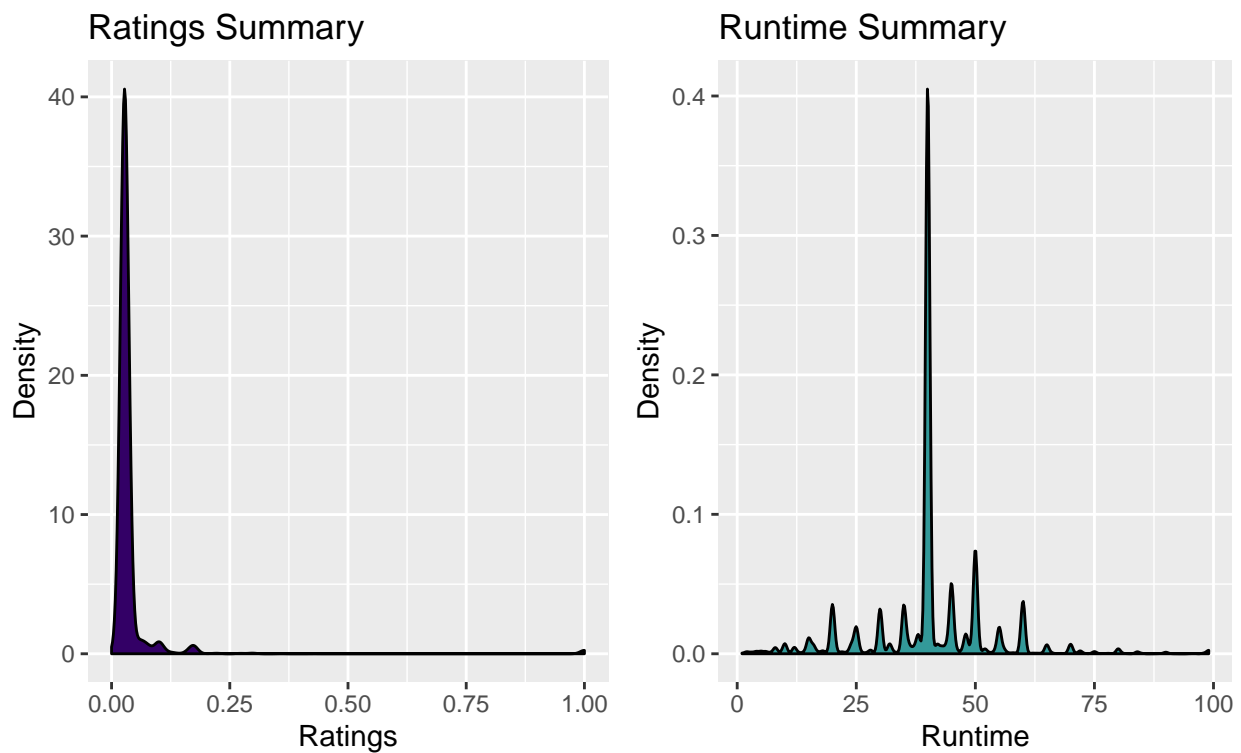
- Categorical: Relationship status, Industry, Genre, Targeted sex, Airtime, Air location and Money back guarantee
- Ordinal: Expensive
- Continuous: Average runtime, Ratings

## Univariate Analysis

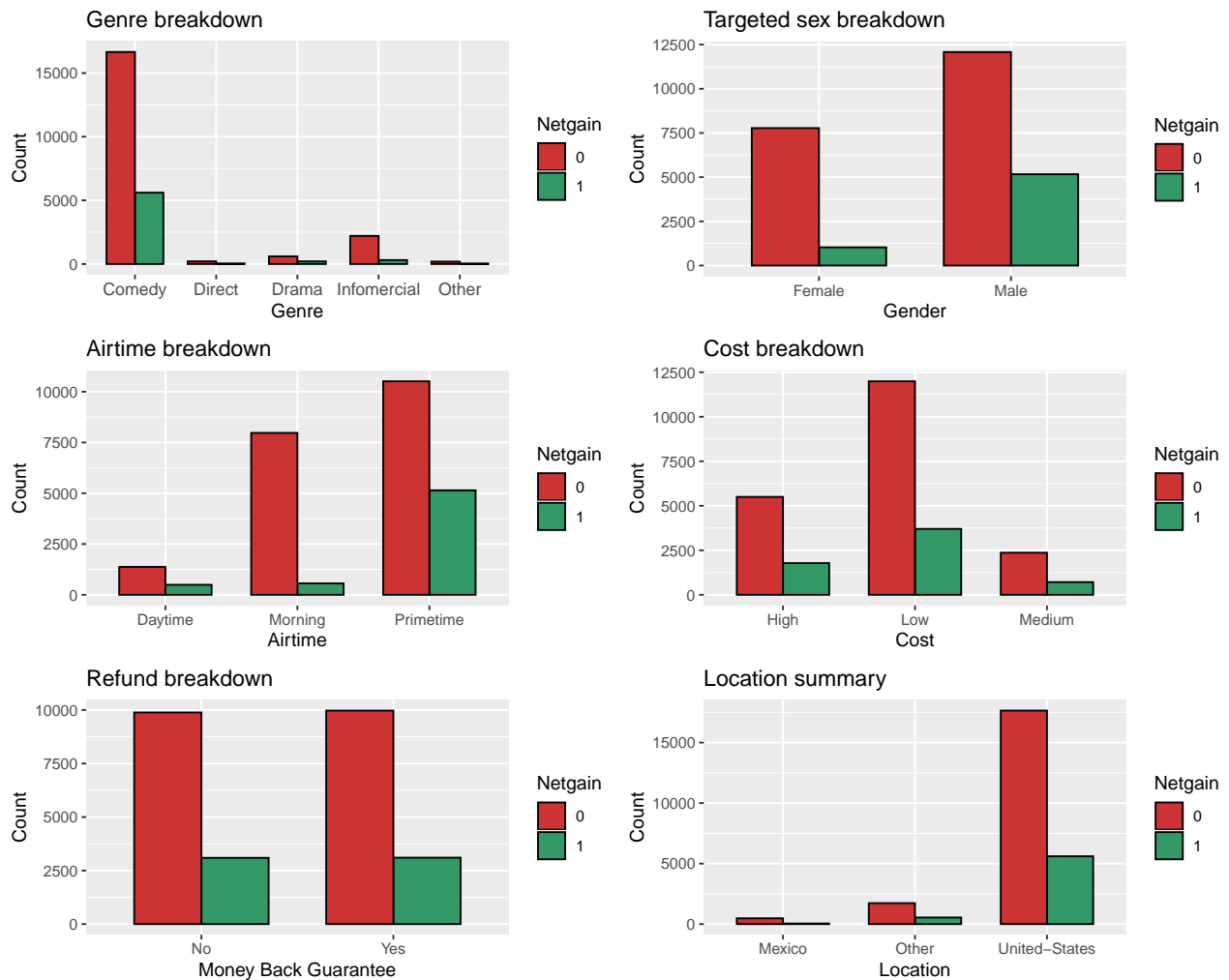
We can look at one variable and vary the colour palette. For this visual, I used the viridis default colour palette.



We can also combine the representations into an array and vary the variables by colour.

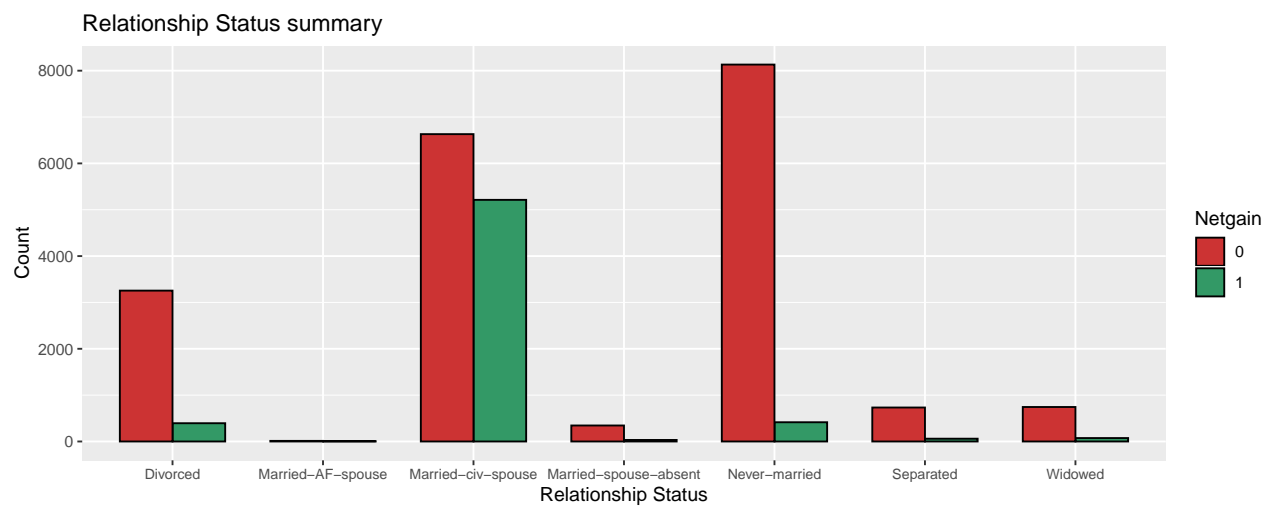
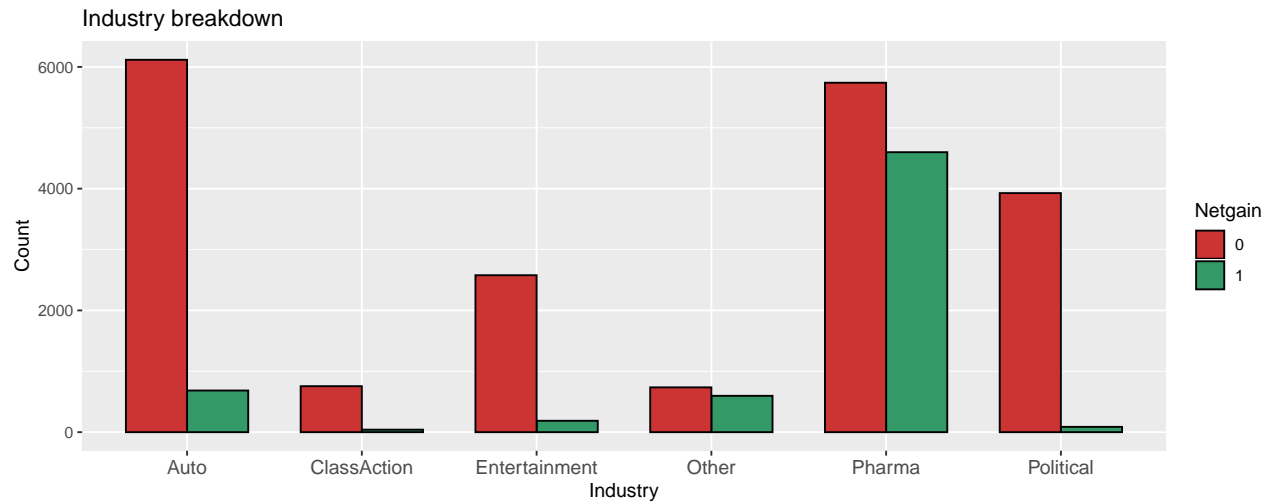


## Bivariate Analysis



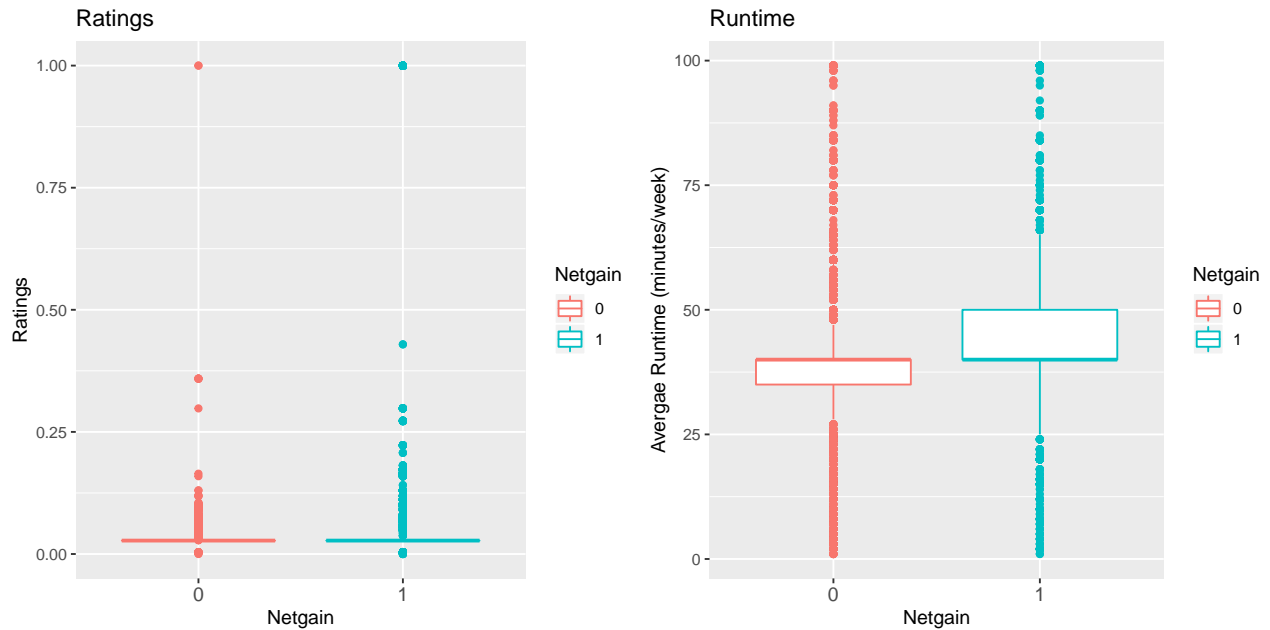
### Analysis

- Genre: We can see that most Ads were comical. While the proportion of dramatic ads seems to be extremely small compared to comedy, the ratio of success seems to be approximately equal.
- Targeted sex: It is quite evident that a higher proportion of ads targted towards men were successful as compared to women.
- Airtime: Ads aired at Primetime seem to have a higher rate of success.
- Cost: While a large proportion of ads were inexpensive, low cost and high cost ads seemed to have an equal success rate.
- Money back guarantee: this seemed to have negligible effect on the success of an ad.
- Location: Most ads were aired in the US. It's not very apparent whether ads aired in other locations had a higher chance of being successful.



Analysis continued

- Industry: Pharma and other seemed to have the highest success rate. The political and class action ads seemed to have the lowest success rate.
- Relationship Status: Married-civ-spouse seemed to have the highest success rate compared to all other statuses.



Analysis continued

- Ratings: As seen in the univariate analysis of ratings, the data points were highly clustered from 0-0.10. Hence, it's not very clear whether the ratings affect the success rate of the ad or not.
- Average runtime: From the boxplot, we can see that successful ads tend to have a higher average runtime per week as compared to unsuccessful ads. Although there is a large proportion of ads that run for around 40 mins, the difference is still noticeable.

## Looking at correlations

```
##               ratings average_runtime.minutes_per_week.
## ratings               1.00                      0.08
## average_runtime.minutes_per_week. 0.08                      1.00
```

Since we only have 2 continuous variables, the correlation is 0.08. This is pretty low hence there is no risk of multicollinearity. We can keep both variables in the dataset.