

Divorce and its Causes in the United Kingdom

Team M

October 13, 2017

Contents

Introduction	2
Description	2
Methods	2
Descriptive Statistics and Correlation Tests	3
Social and Economic Status & Marriage	3
Education	6
Religion and Culture & Marriage	25
Country of birth	29
Religion	30
Age & Marriage	31
Region/Geography & Marriage	34
Dependent Children and Social Grade	37
Health & Marriage	37
Logistic Regression	39
Hypothesis Testing and Prediction Model (Helene's)	41
Conclusion	41
References	41
Appendix	41

Introduction

Marriage is viewed by many as an important social and sometimes religious life event for a majority of people. It symbolizes and unites two individuals into a single-family unit of everlasting bliss. Tragically, for some marriages face the reality that they were not meant to last and either officially through divorce or unofficially through separation. Since the United Kingdom divorce rate hit its peak in 2003, it has steadily dropped to NUMBER HERE (Haines, 2017) and CITATION NUMBER. This drop has been attributed to culture changes around cohabitation, “couples living together before getting married” (Bingham, 2015).

Despite the wonderful change in the overall divorce rate, there are still many new divorces and separations each year leading us to wonder what could still cause such a travesty. Our curiosity and ultimately investigation is to find the remaining factors still bringing about the decline and end of family units across the United Kingdom based on the 2011 Census data.

Description

The dataset we examined is the 5% sample data of the 2011 UK Census. It contains 2,848,155 observations and 121 variables. Each observation represents a UK citizen. Since the dataset is very large and there are too many variables, we decided to exclusively focus on the relationship between marital status (`marstat`) and other variables. In brief, we would like to examine the factors that may influence people’s marriage and lead to higher divorce rate. Therefore, we selected a total of 17 variables that may be highly correlated with marital status.

We can attach a table of all variables we examined here:

Methods

a really brief description of the R packages we used. We can load all our libraries in this page:

```
library(dplyr)
library(ggplot2)
library(gridExtra)
library(stargazer)
library(readr)
library(RCurl)
library(foreign)
```

In our report we aim to explore how selected variables singularly and collectively affect the probability of an individual being divorced or separated.

In the first part of the report we explore the interactions between selected variables and divorce rate. The divorce rate is calculated as percentage of unhappy marriages over total marriages,

$$(divorced + separated) / (married + samesex + divorced + separated)$$

We include separated in our definition of unhappy marriage because sometimes people continue to stay in marriage for practical reasons such as cohabitating and tax benefits, but that is more form over substance and most separated couples end up in divorce. Our team is particularly interested in the factors that lower people’s marriage satisfaction which eventually lead to divorce.

The variables we select include inherent characteristics of an individual such as age, ethnicity, religion and country of birth, characteristics an individual acquire such as social grade, highest qualification achieved, and disability and lastly external factors such as region and type of housing.

Exploratory statistics includes mostly plotting. Bar plots are used when we want to represent the divorce rates for different classes of each variable, for example highest qualification achieved ranges from no qualification to level 4 (university and above). When we want to explore the interactions between more than one variables, we use color to represent them on the graph or use a grid to lay multiple small scale plots next to each other in order to identify any difference in pattern.

Hypothesis testing is also used to test if the observed pattern is statistically significant. Chi-squared test is used to explore whether two categorical variables are independent. Correlation test is used for testing correlations between two samples. Since most of our variables are categorical and marital status is categorical, chi-squared test is the best choice in our analysis. T-test is used for testing of difference in divorce rates across classes of variables. We use 5% significance level for all the hypothesis testing done in this research. If the p-value from the test is less than 5%, we can confidently reject the null hypothesis.

Descriptive Statistics and Correlation Tests

Simple descriptive statistics including bar charts, pie charts, density graph... etc. We can show a few correlation tests, chi-square tests and graphs

Social and Economic Status & Marriage

```
## [1] 955170      5
```

After filtering out people whose education level cannot be interpreted and filtering out missing values 1.684.636 observations are left. The distribution between social grades is alike. Education on the other hand has a significant underrepresentation of participants that absolved an apprenticeship. They account for less than 4 % of the surveyed participants. This should be kept in mind when interpreting divorce rates in the apprenticeship group.

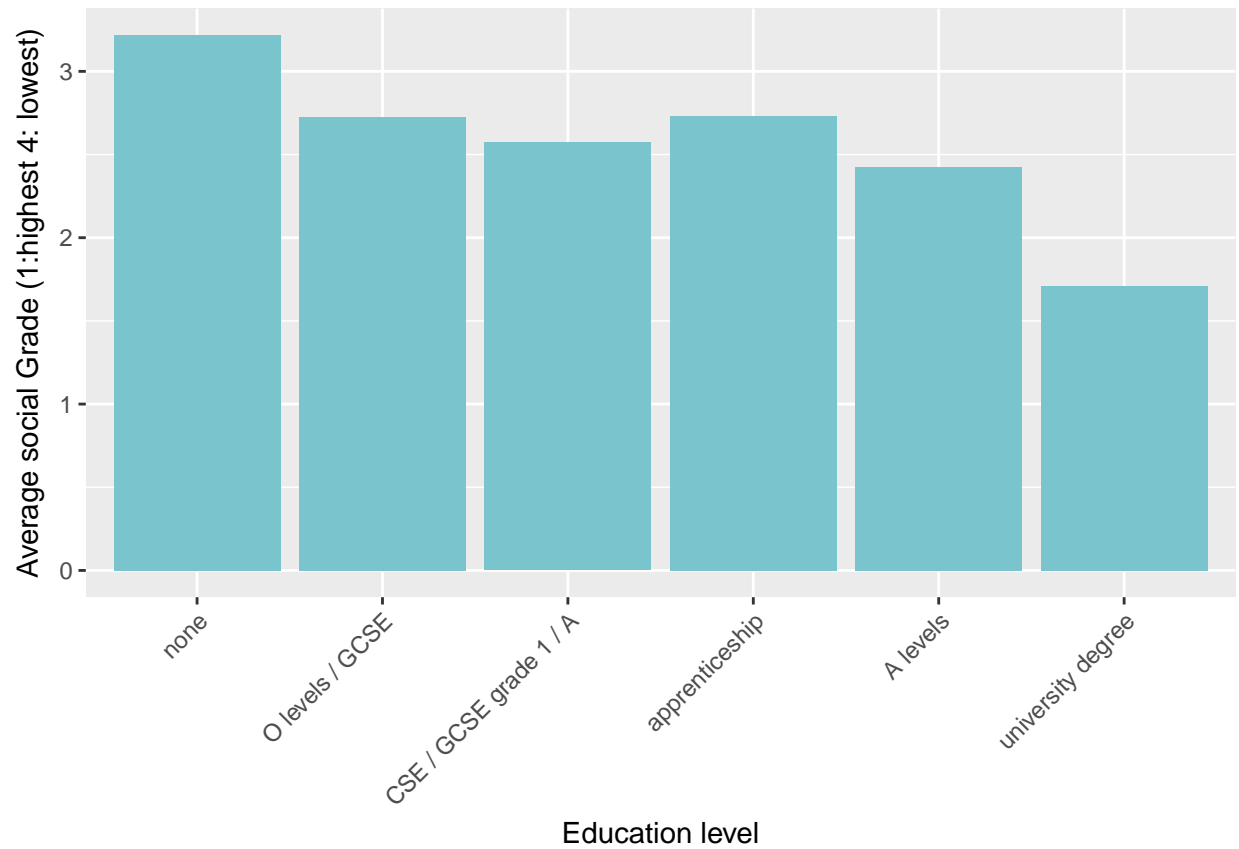
Between education and social grade is a middle strong linear correlation. The higher the education level the higher the social grade. Still it is insightful to look at both variables alone and combined, as they give different insights.

```
##
##      0      1
## 749952 205218

##
##      0      1      Sum
## 0.7851503 0.2148497 1.0000000

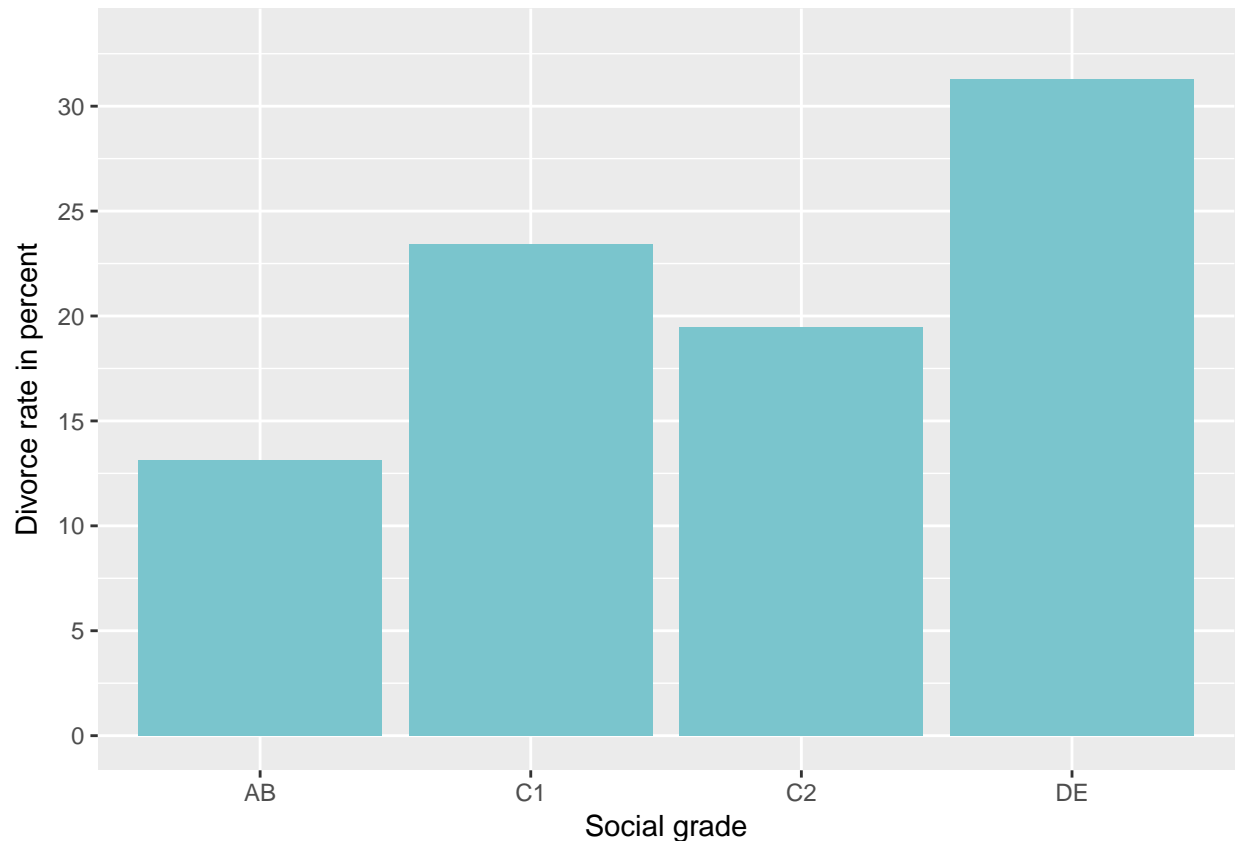
##
##      10      11      12      13      14      15
## 162768 150725 153917  37220 117626 332914

##
##      10      11      12      13      14      15
## 0.17040736 0.15779914 0.16114095 0.03896689 0.12314666 0.34853900
```



```
## # A tibble: 6 x 3
##   education      n mean_grade
##   <int> <int>     <dbl>
## 1      10 162768  3.216314
## 2      11 150725  2.725858
## 3      12 153917  2.572179
## 4      13  37220  2.733557
## 5      14 117626  2.425510
## 6      15 332914  1.708339

## [1] -0.4973118
```



For social grade the finding match the underlying hypothesis fundamentally. Social grade is directly connected to living standards. Society holds the common belief that money is the root of all marriage problems. Nevertheless, there are a few studies supporting this theory and many contradicting it (Andersen, 2008). http://www.tandfonline.com/doi/abs/10.1300/J087v43n01_08 According to the UK census data, this assumption seems to prove true: the lowest social grade has an above average divorce rate of 33 % versus 13 % for the upper and middle class. Only between social grade C1 and C2 the pattern does not apply. C2 has a 4% lower divorce rate compared to C1. But still, both lie in the middle of AB and DE and are close to the total average divorce rate of 21 %.

To see if the social grade is a true indicator for divorce and not only related to the fact, that migrants, different religions or higher educated, who tend to have a lower divorce rate than the British, are overrepresented in some social groups, the data was crosschecked with religion as well as ethnicity.

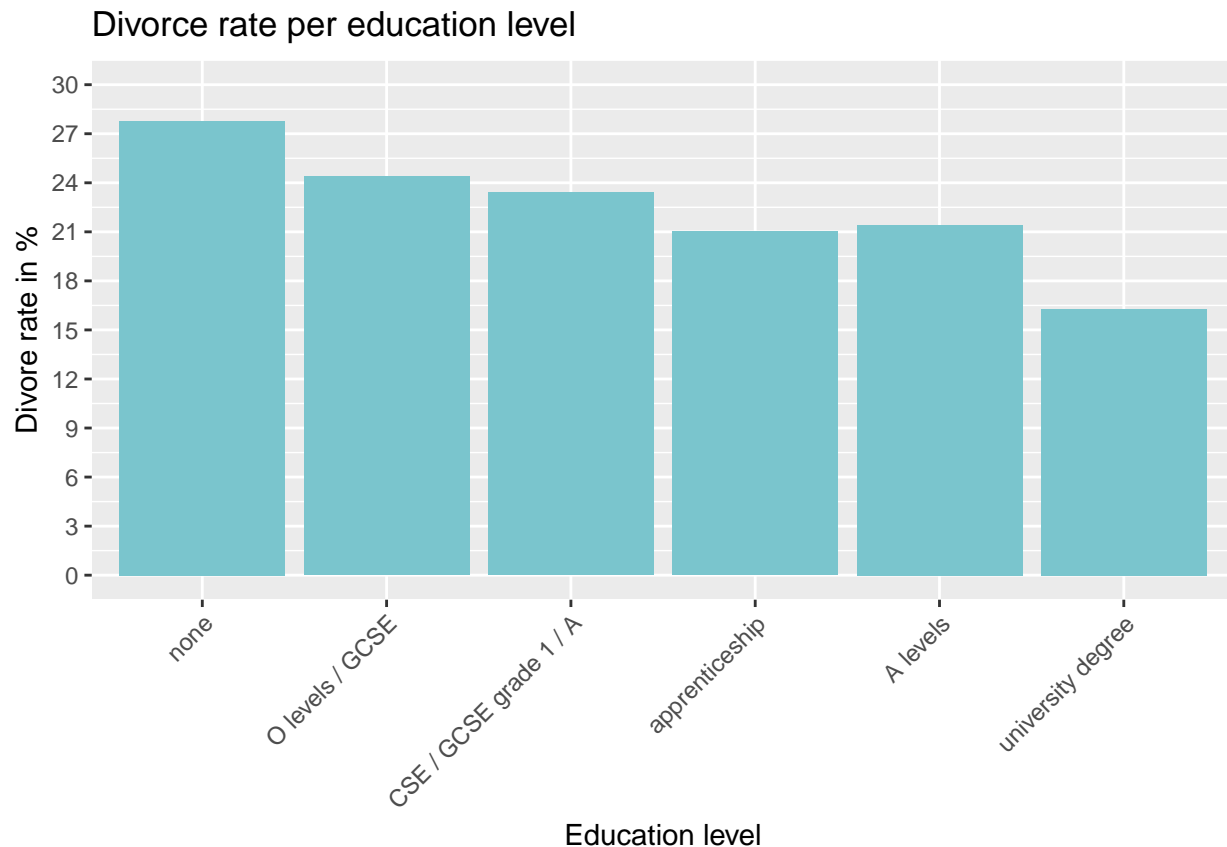
The exact same pattern, also with C1 being slightly higher than C2 holds across all religions, ethnicities and education levels. Therefore it can be concluded, that social grade has a direct influence on the divorce rate. Upper and middle class have a way lower probability of getting divorced than all other groups. A meaningful factor is the higher income, which enables better housing, better healthcare, less stressful life, more security and therefore less severe problems affecting the average marriage. Next to monetary problems, lower social groups tend to have a higher probability of drug and alcohol problem, one of the major reasons for divorce (NHS 2014; Trail and Karney, 2012). Additionally, earlier marriages, which will be explained in the education evaluation, are a burden for marriages.

<http://digital.nhs.uk/catalogue/PUB15943> <http://onlinelibrary.wiley.com/doi/10.1111/j.1741-3737.2012.00977.x/full>

Across all sociodemographic characteristics (religion, ethnicity, education levels) the lowest social grade always has a significantly higher divorce rate and the highest social grade the lowest divorce rate. Overall it can be concluded that social grade, which is highly correlated with income, is a strong driver for divorces. The

difference between social grade C1 and C2, which is as well constant across different sociodemographic groups, is a question for further research and could not be evaluated based on the census data.

Education



The graph clearly confirms our expetations: the higher educated the lower the divorce rate. As can be seen, the difference in divorce rate between no degree and a university degree is about 10%. People with no education have a divorce rate of 28 %, univeristy graduates 17 %. People with a school education approach the mean divorce rate of 21 % the higher their school degree is.

A reason for the negative correlation directly linked to education, is that higher educated people marry later than less educated people. A reason for marrying later the higher educated someone is, is that the education lasts longer and therefore “adult life” and a stable lifestyle starts later. More educated people tend to focus more on their career in their jounge ages. Especially for women, a higher education results in a higher independence which lowers her feeling of need for getting married early (Isen, 2010). Additionally, higher educated people are in a relationship and live together longer time before getting married. Knowledge about the partner, expectations and security if both share the same ideas for the future grow with age and experienced with each other. <https://www.theguardian.com/lifeandstyle/2009/sep/27/divorce-linked-to-age-education> <http://www.nber.org/papers/w15725.pdf> Furthermore, being pregnant or having children is still a reason for early marriage which results in a higher divorce rate. As less educated women tend to have children when their younger, they marry earlier because of the child (Office for National Statistics, 2015).

A result from the higher education is a better financial situation which lowerst the divorce rate, as seen in the social grade analysis. As 80% of the people marry in their own social category, the gap between income and education between different couples grow even stronger.

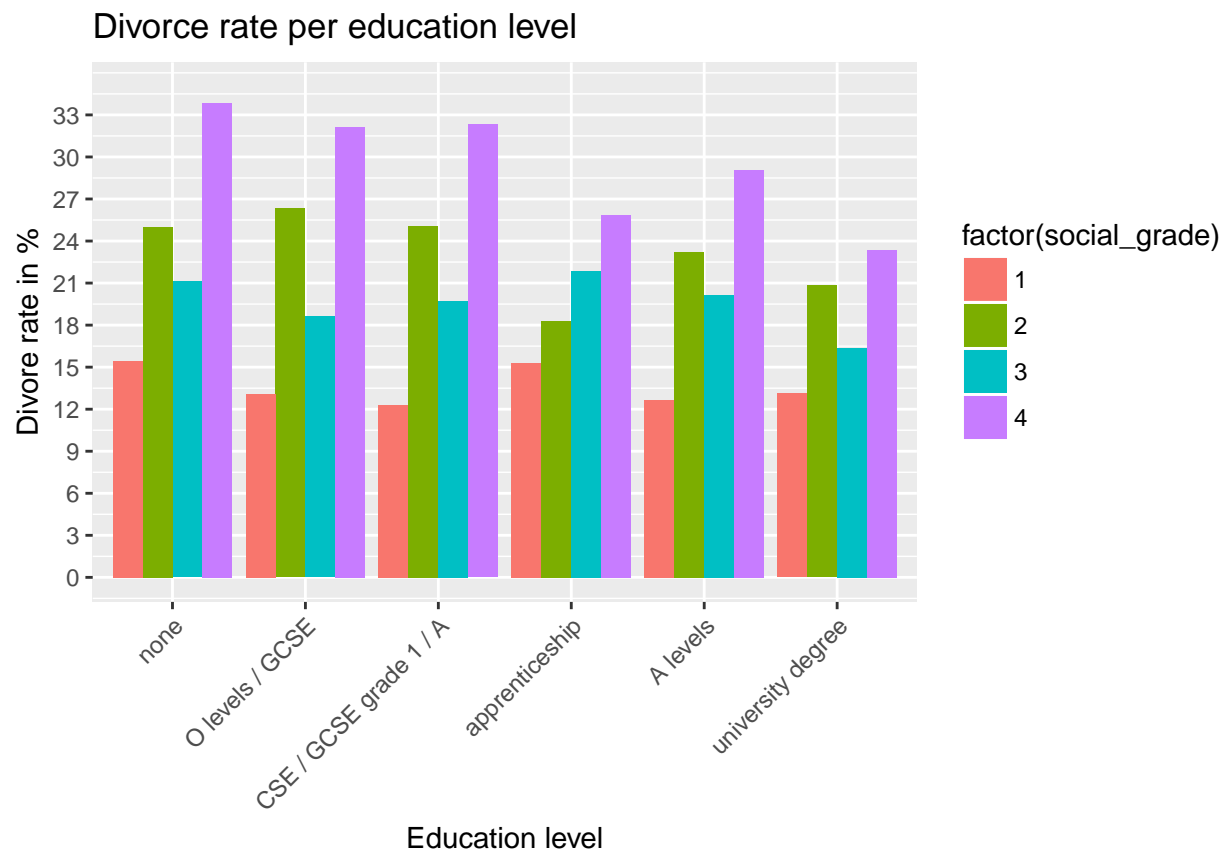
<http://www.sueddeutsche.de/wirtschaft/partnerwahl-und-ungleichheit-heiraten-zementiert-soziale-spaltung-1.1875581> (<http://www.pewresearch.org/fact-tank/2015/12/04/education-and-marriage/>)

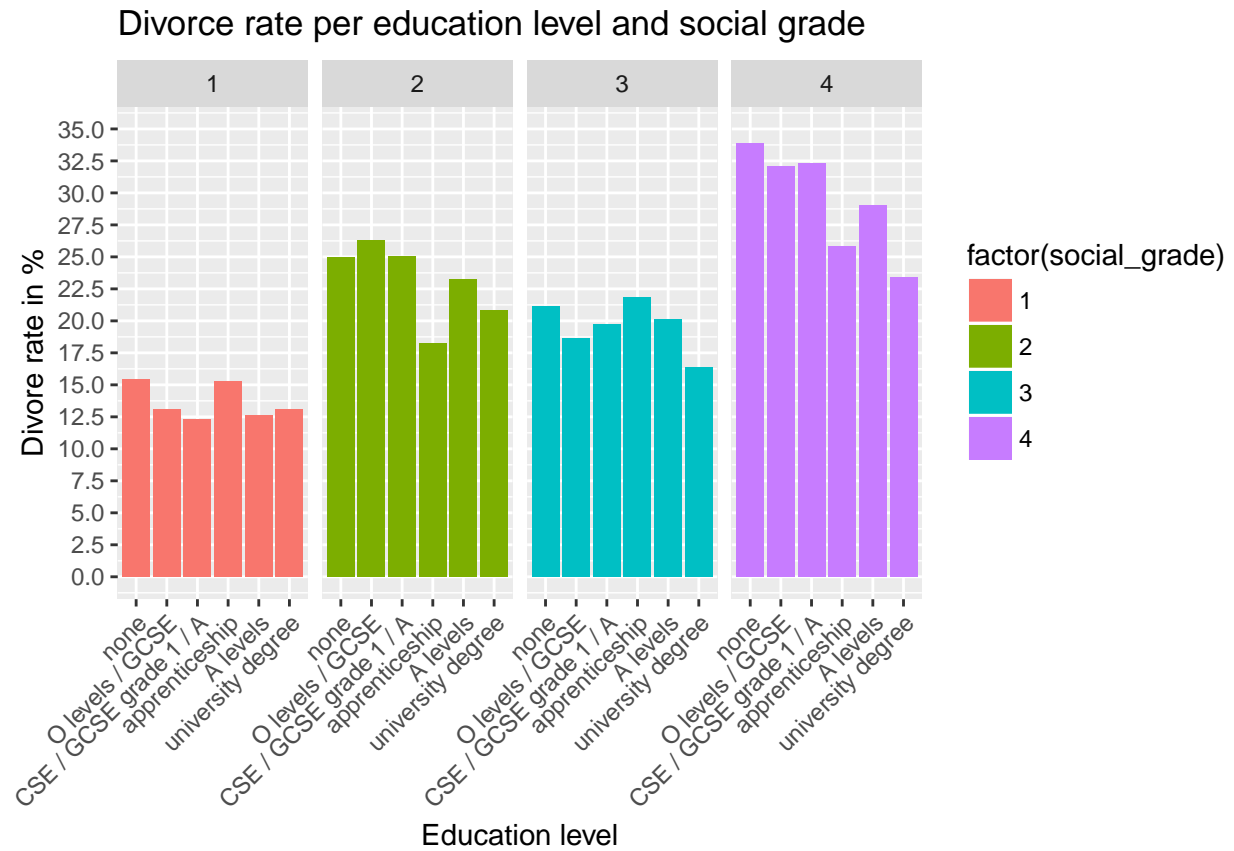
<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsbyparentscharacteristicsinenglandandwales/2015>

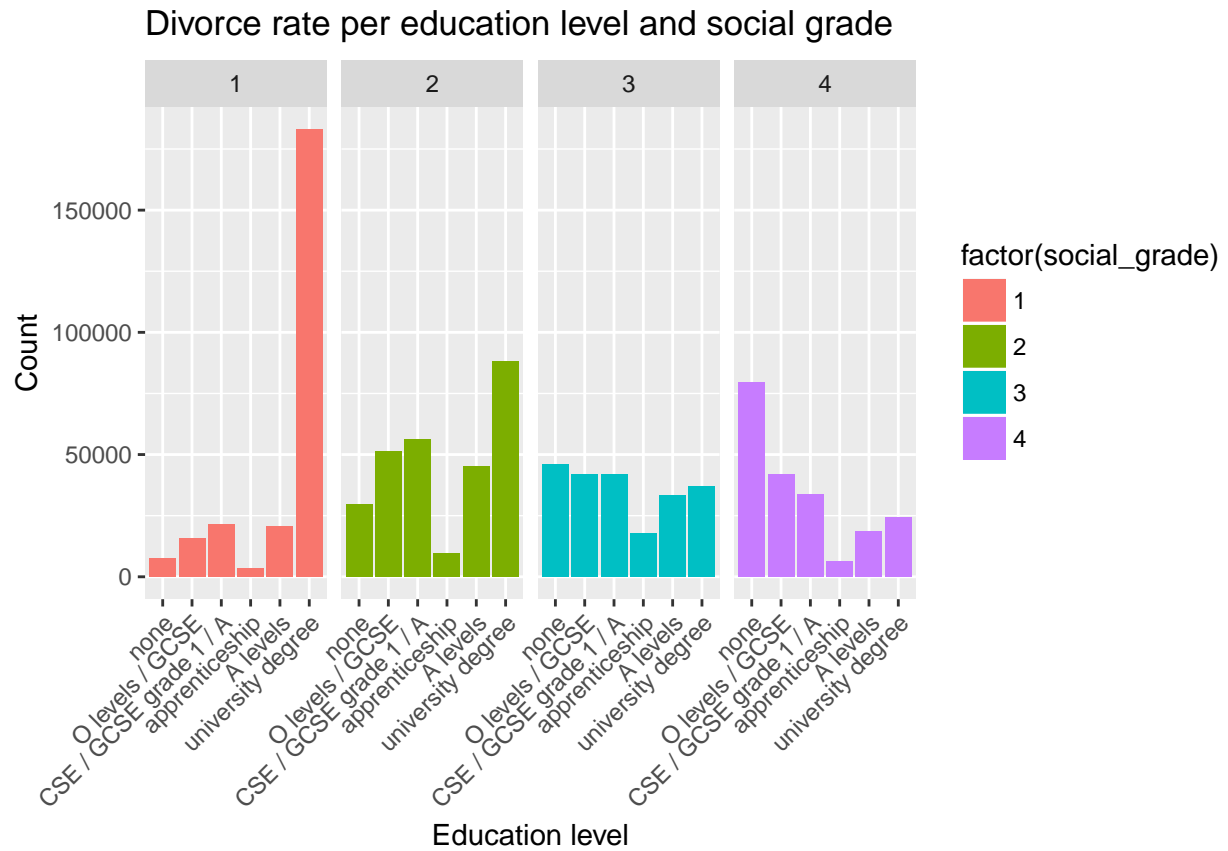
By looking at the relation between divorce rate and education across different religions and ethnic groups, it can be seen that the negative correlation holds. Education therefore seems to be a driver for divorce independent of the religion and ethnicity. This strong pattern diminishes when looking at the highest degree in every social group. The negative correlation is still present, but as education is highly inequally distributed among social grades, social grade acts as a confounding variable. The strong pattern of a lower divorce rate in the social group AB is partly driven by the majority of the people belonging to this social class being highly educated

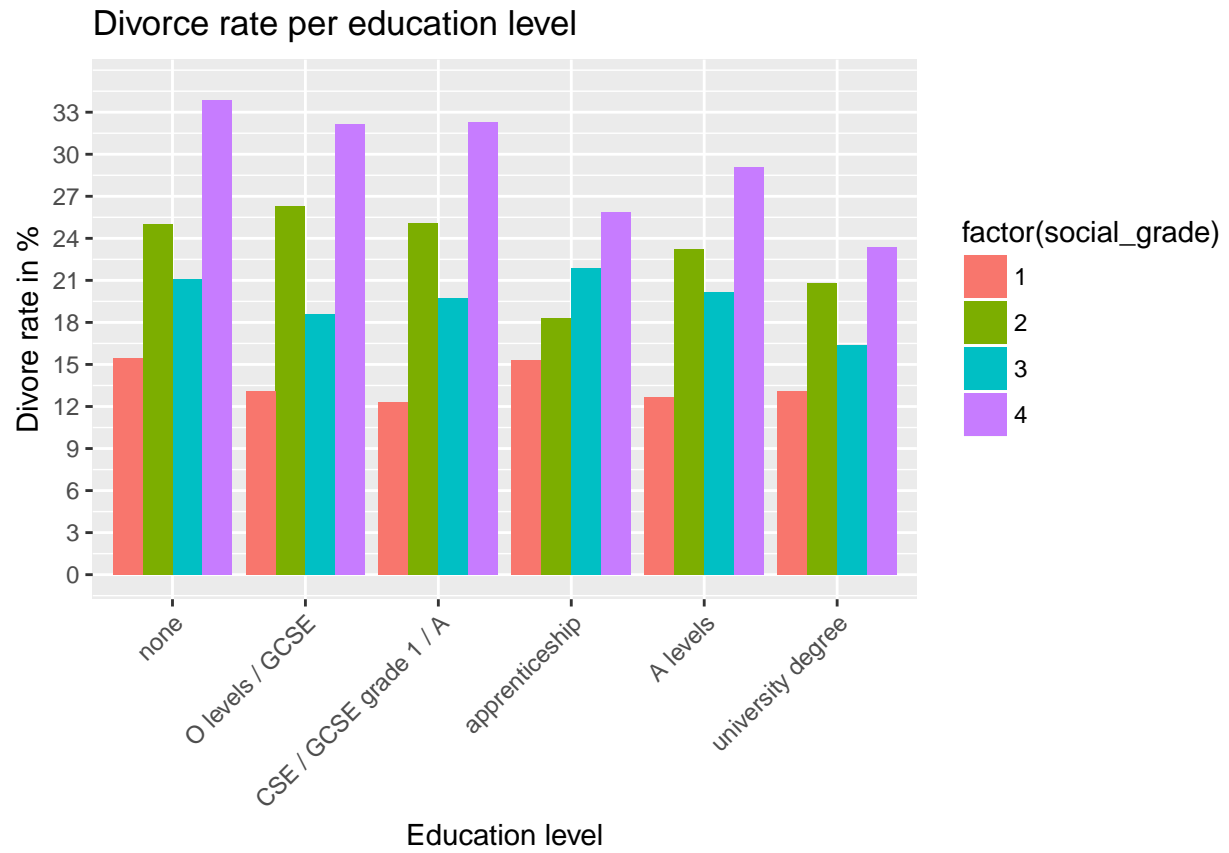
It can therefore be concluded that both, education and social grade, are strong driver of divorce. The unequally distributed education levels within social classes reinforce this effect. The highest overall divorce rate of 34% is thus in the lowest social grade group for people with no education at all. This effect is reinforced for black people in the lowest social grade with no education. Their divorce rate is 37% on average. With a divorce rate of only 6% highly educated Asians in the AB social group hold the lowest divorce rate.

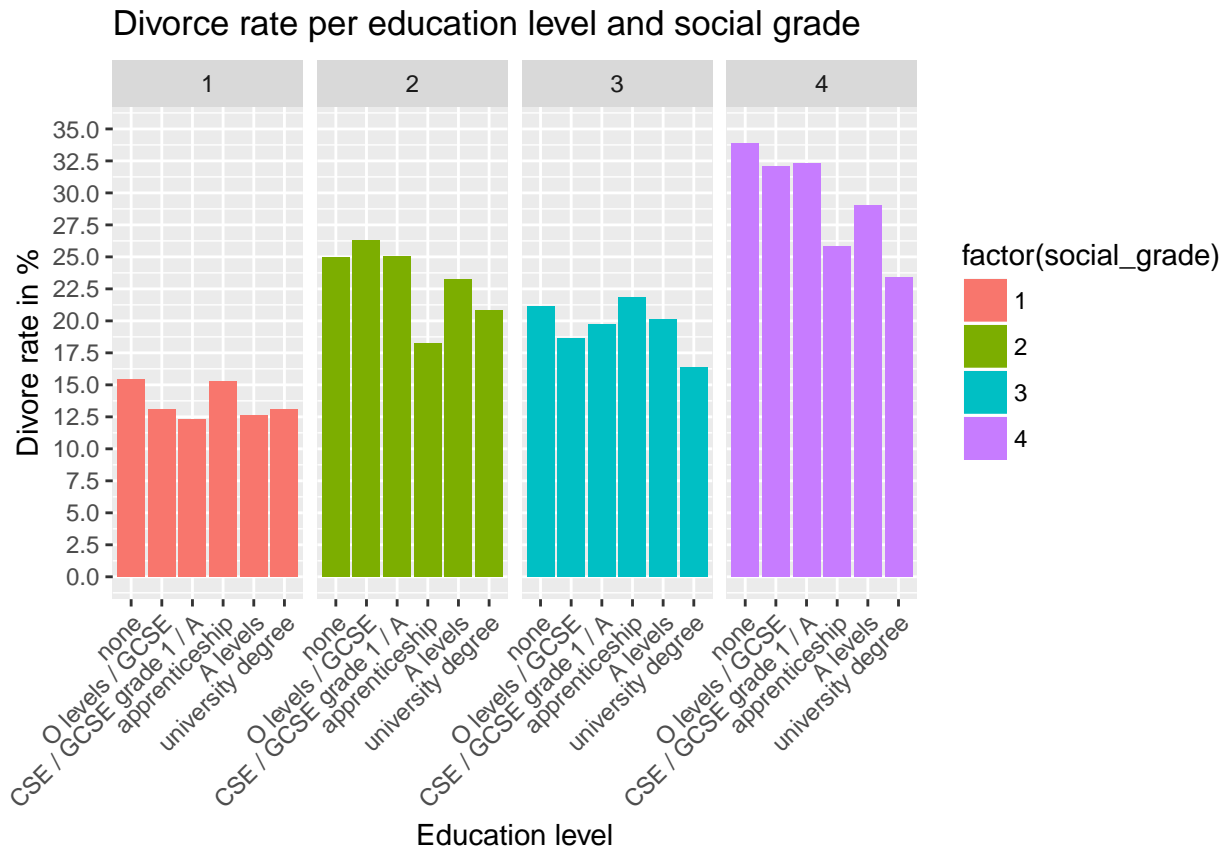
The average british person is white, christian, belongs to the middle class and has a university degree has a average divorce rate of 12 %. By only changing education (but keeping social class) the divorce rate ascends to 14.5 %. When looking for both education extremes in the lowest social class we get a divorce rate for university graduates of 24 % versus 35 % for non educated people. It can be seen by this, that social grade accounts for higher differences in divorce rate than education does.











```
## [1] 955170      5
```

For education level “other” is a foreign degree and therefore not applicable and excluded. Attenteion when we look at immigration backround, as those people might be crossing.

Social grade: 1 AB (highest) 2 C1 3 C2 4 DE (lowest) -9 were excluded as it represents “persons resident in a communal establishment OR persons aged under 16 OR persons aged over 64 OR student/schoolchild living away during term-time OR short-term migrant” (Attention when we match with other data which might include people over 64)

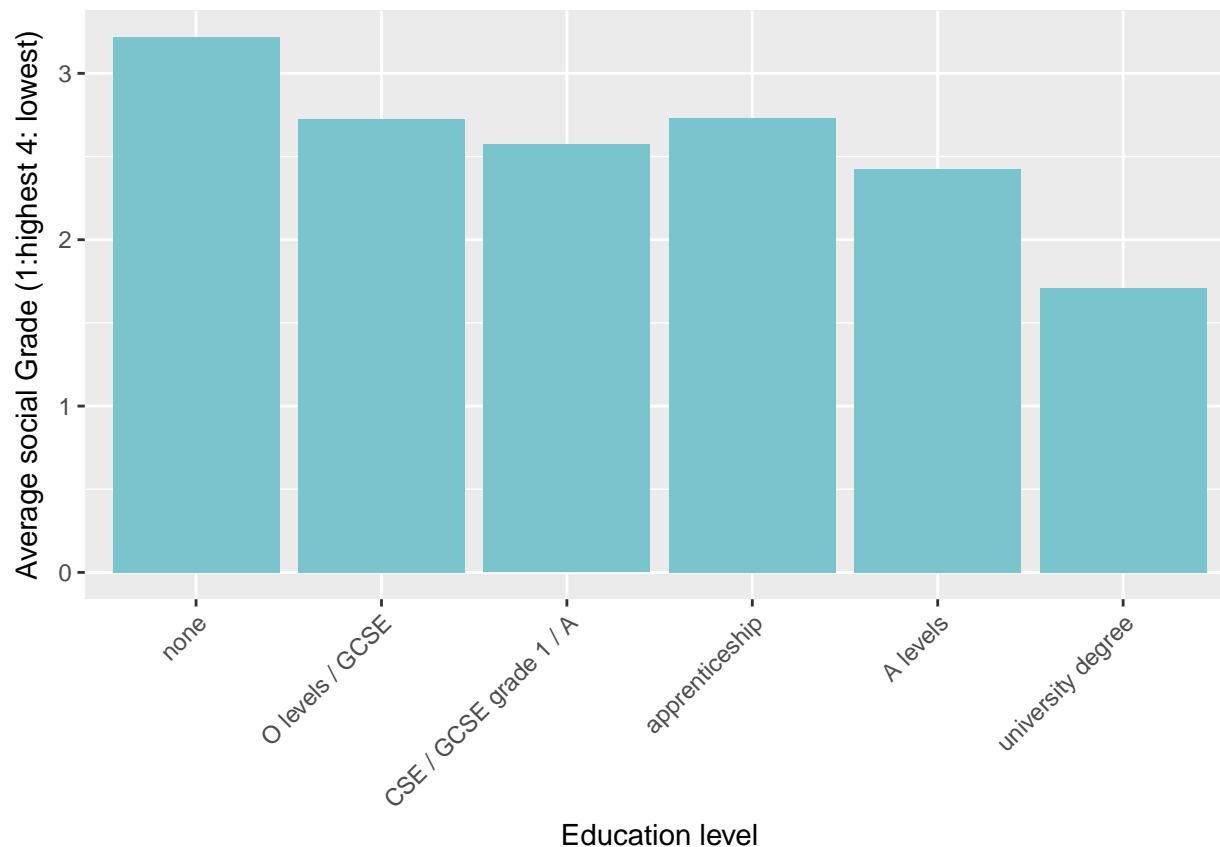
After filtering out people whose education level cannot be interpreted and filtering out missing values 1.684.636 observations are left.

```
##
##      0      1
## 749952 205218

##
##      0      1      Sum
## 0.7851503 0.2148497 1.0000000

##
##      10      11      12      13      14      15
## 162768 150725 153917 37220 117626 332914

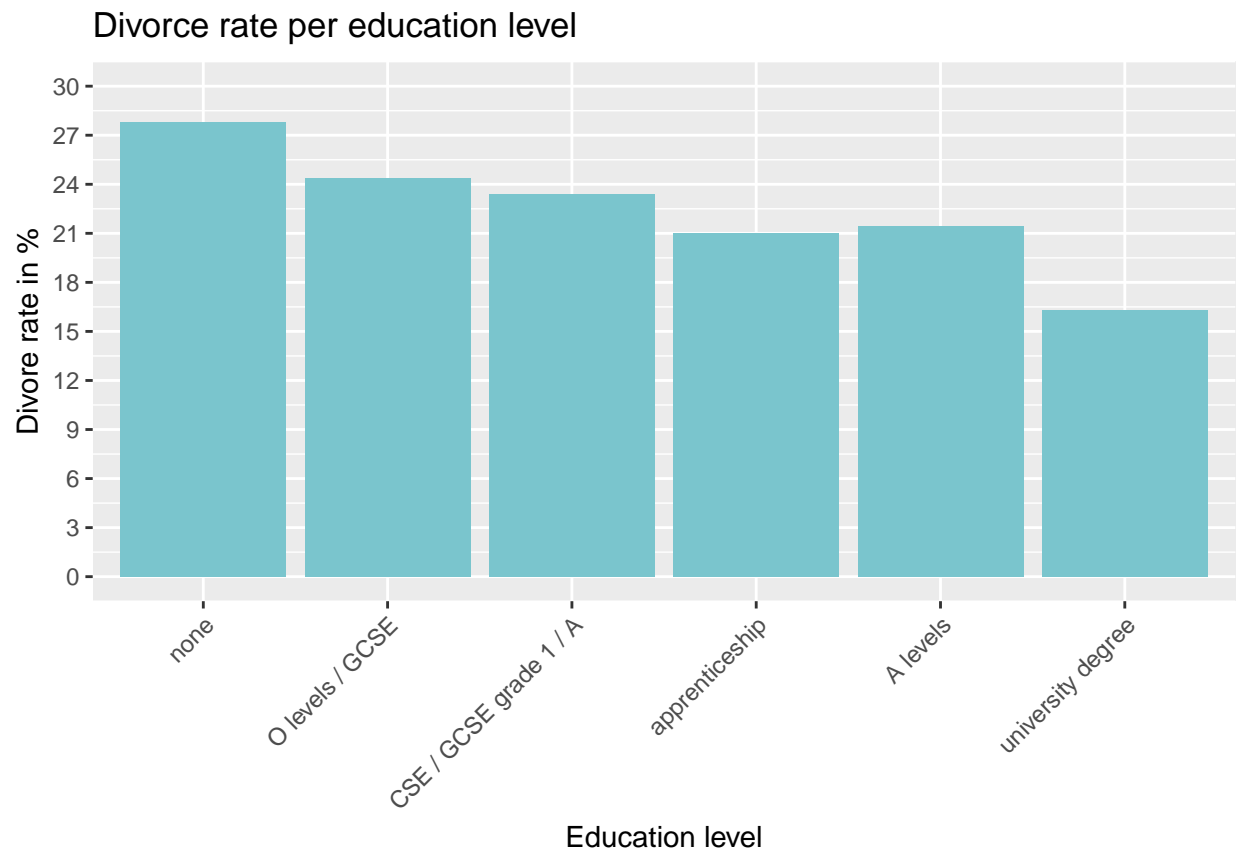
##
##      1      2      3      4
## 252090 280110 218144 204826
```

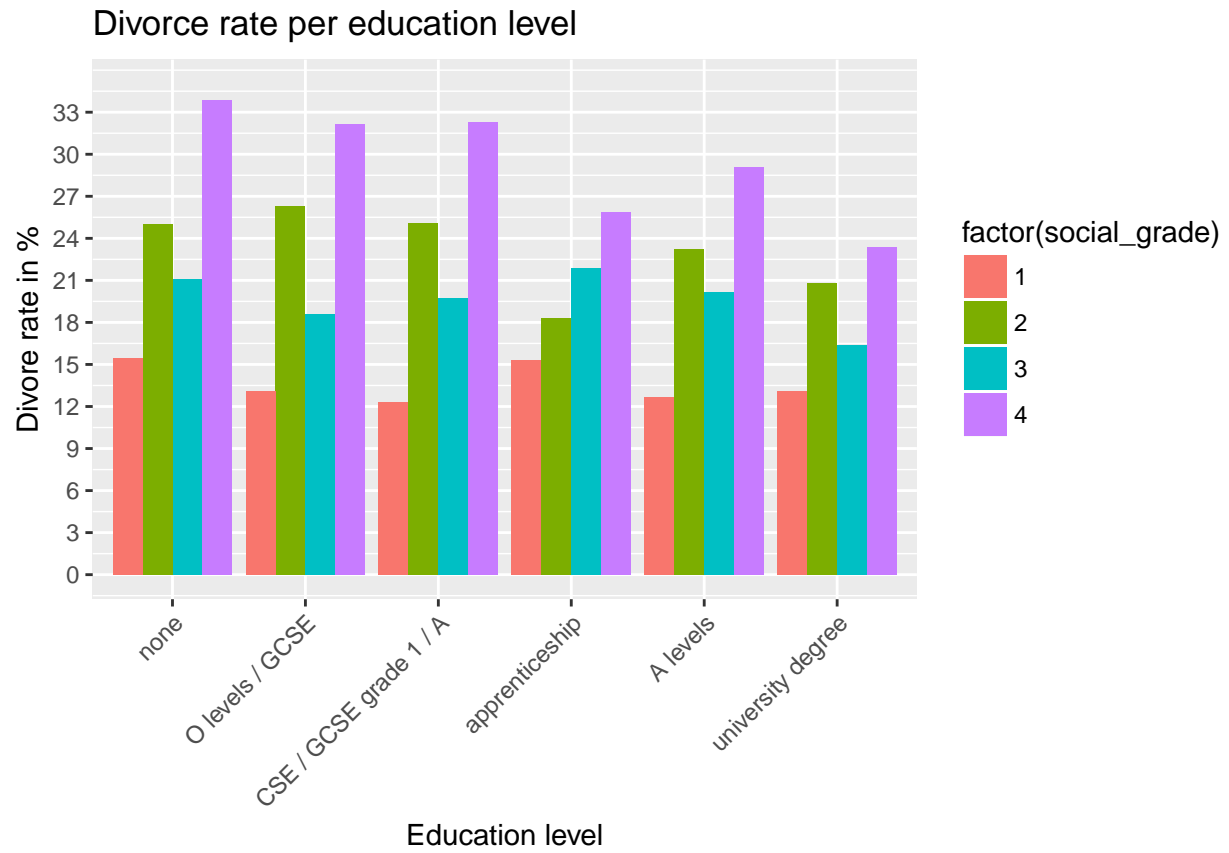


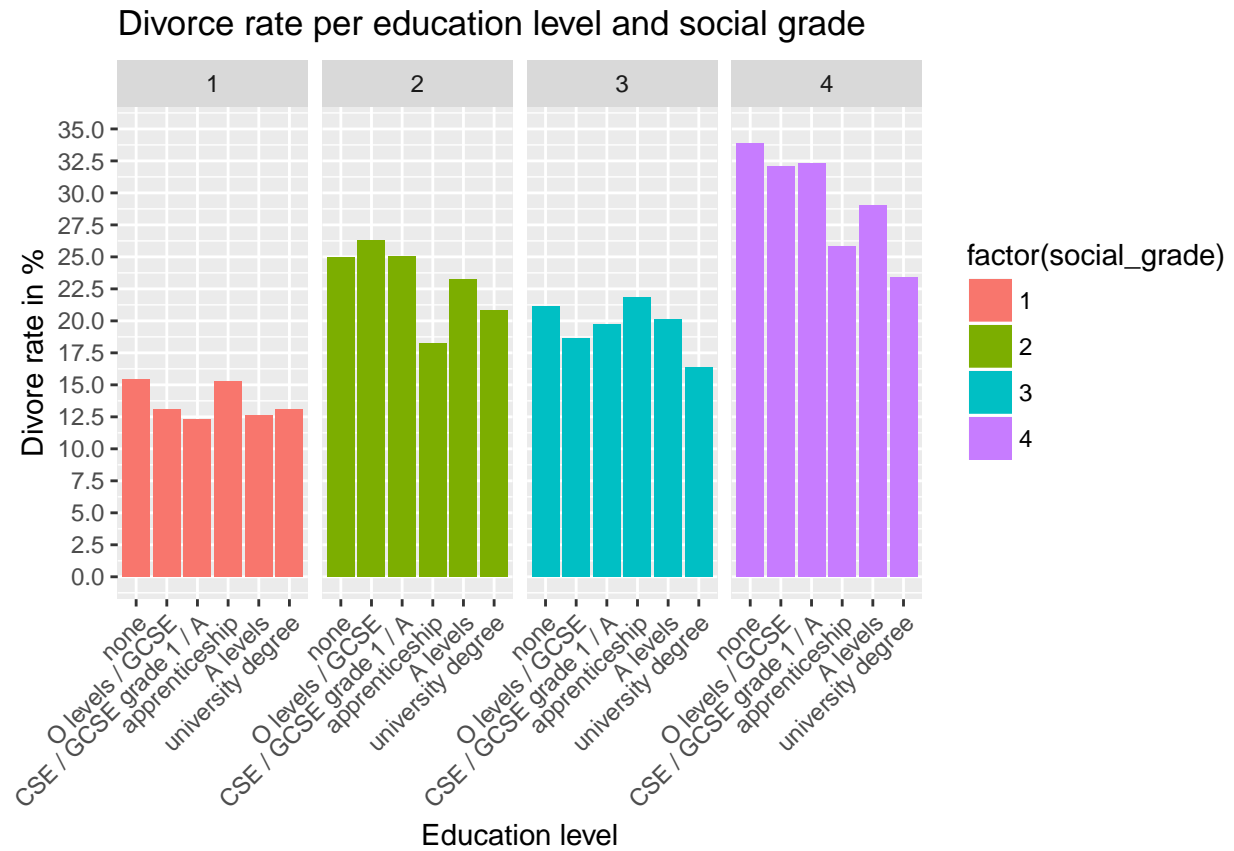
```
## # A tibble: 6 x 3
##   education      n mean_grade
##   <int> <int>    <dbl>
## 1      10 162768  3.216314
## 2      11 150725  2.725858
## 3      12 153917  2.572179
## 4      13  37220  2.733557
## 5      14 117626  2.425510
## 6      15 332914  1.708339

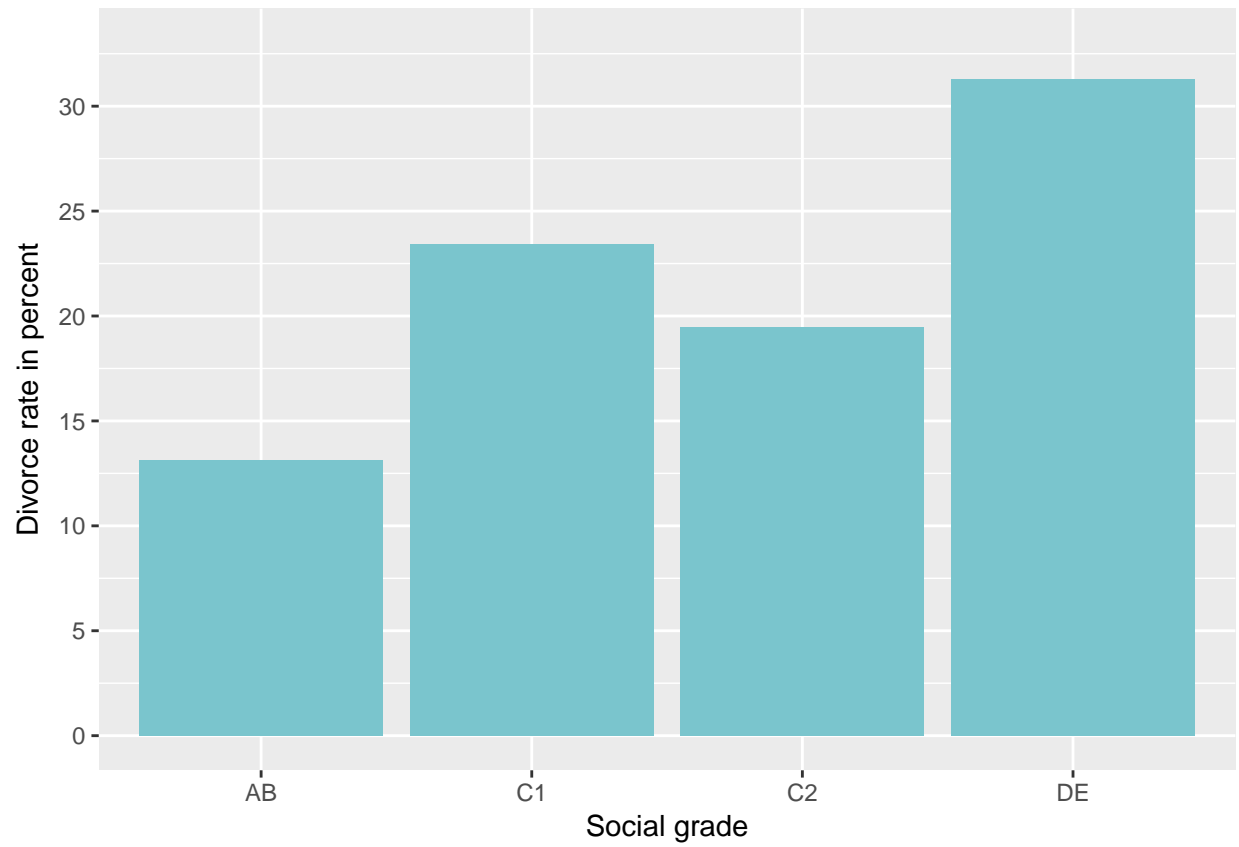
## [1] -0.4973118
```

The data shows, that the higher the education, the higher the social grade. Exept for people who did an apprenticeship. they have a slightly lower social grade than people with CSE/GCSE. There is a middle strong linear correlation between social graed and education level. The higher the education level the higher the social grade.



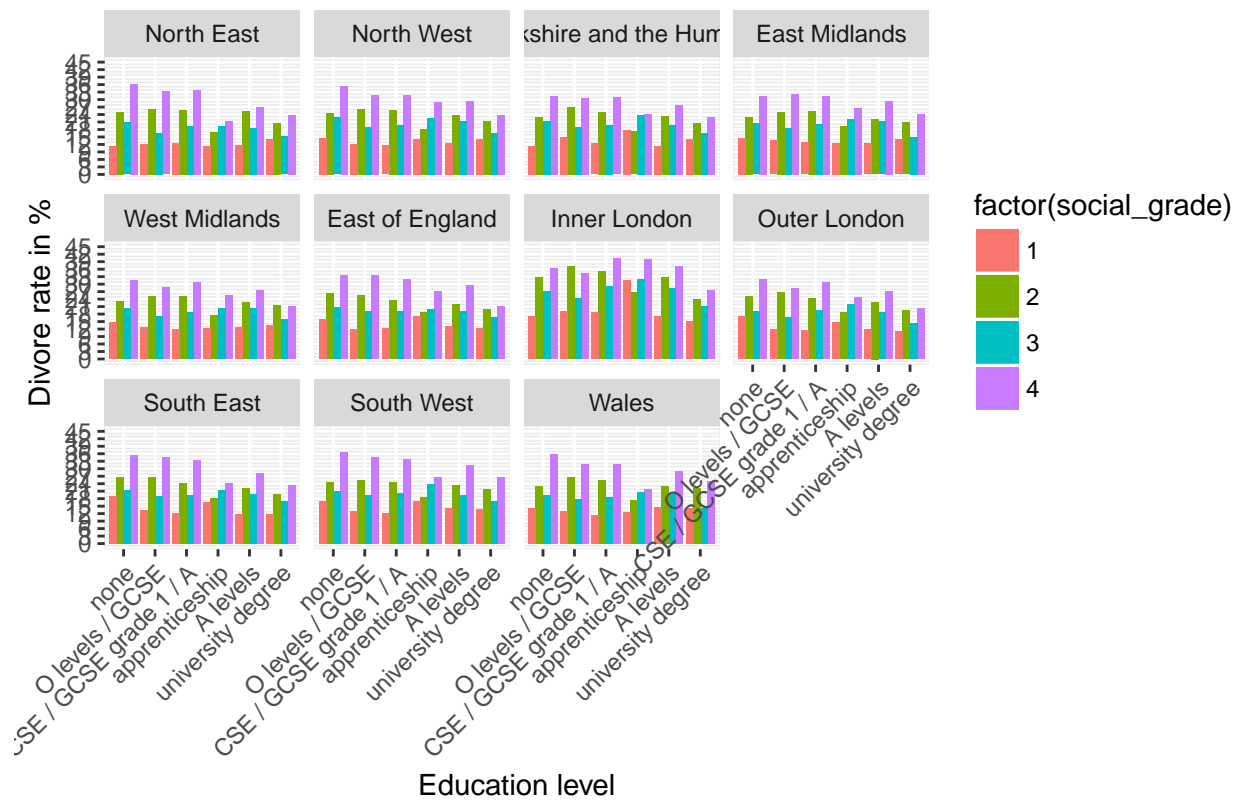






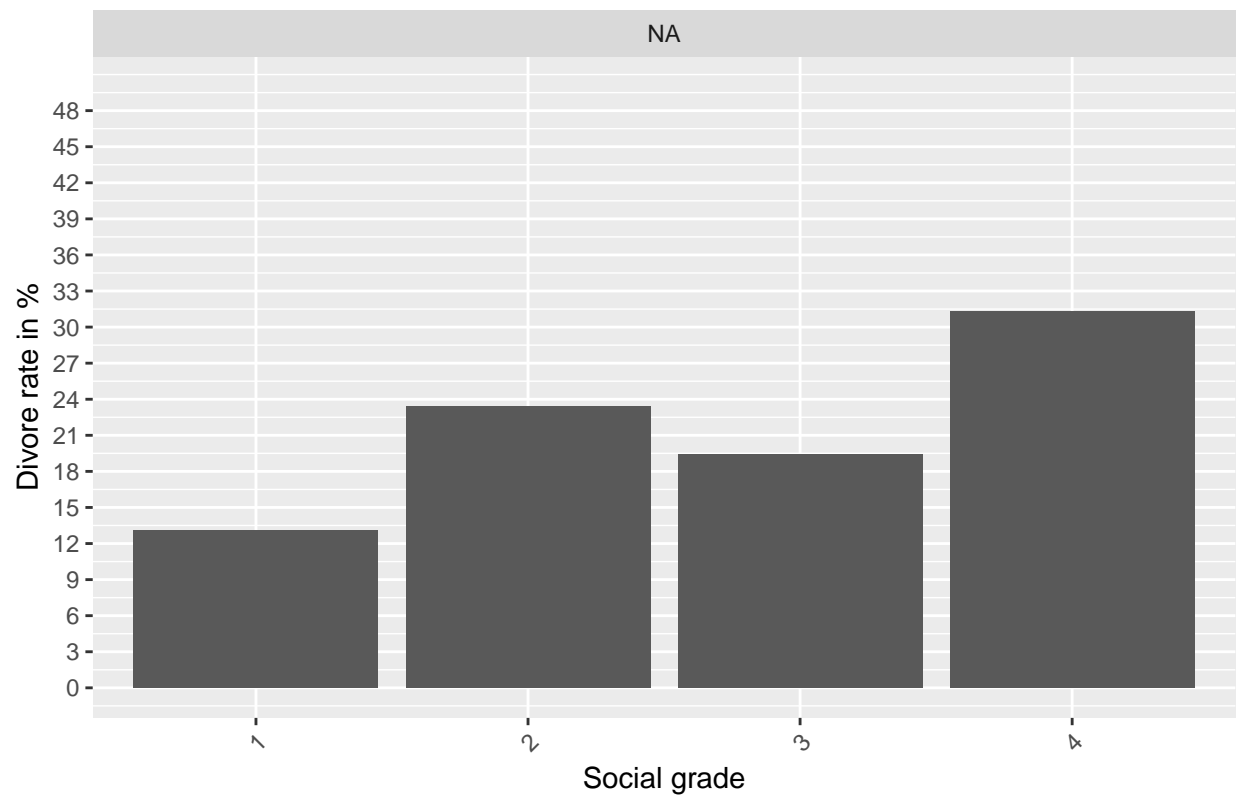
```
## [1] 955170      5
```


Divorce rate per education level and region

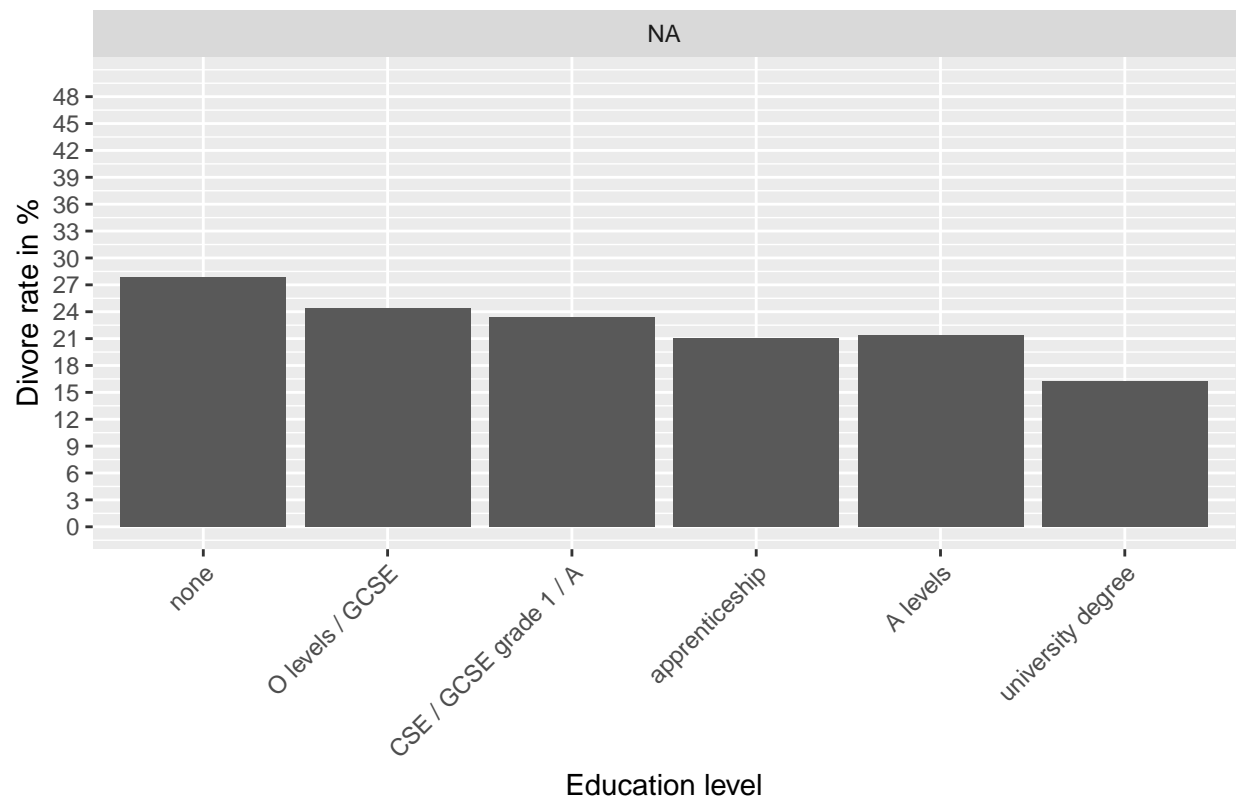


[1] 955170 7

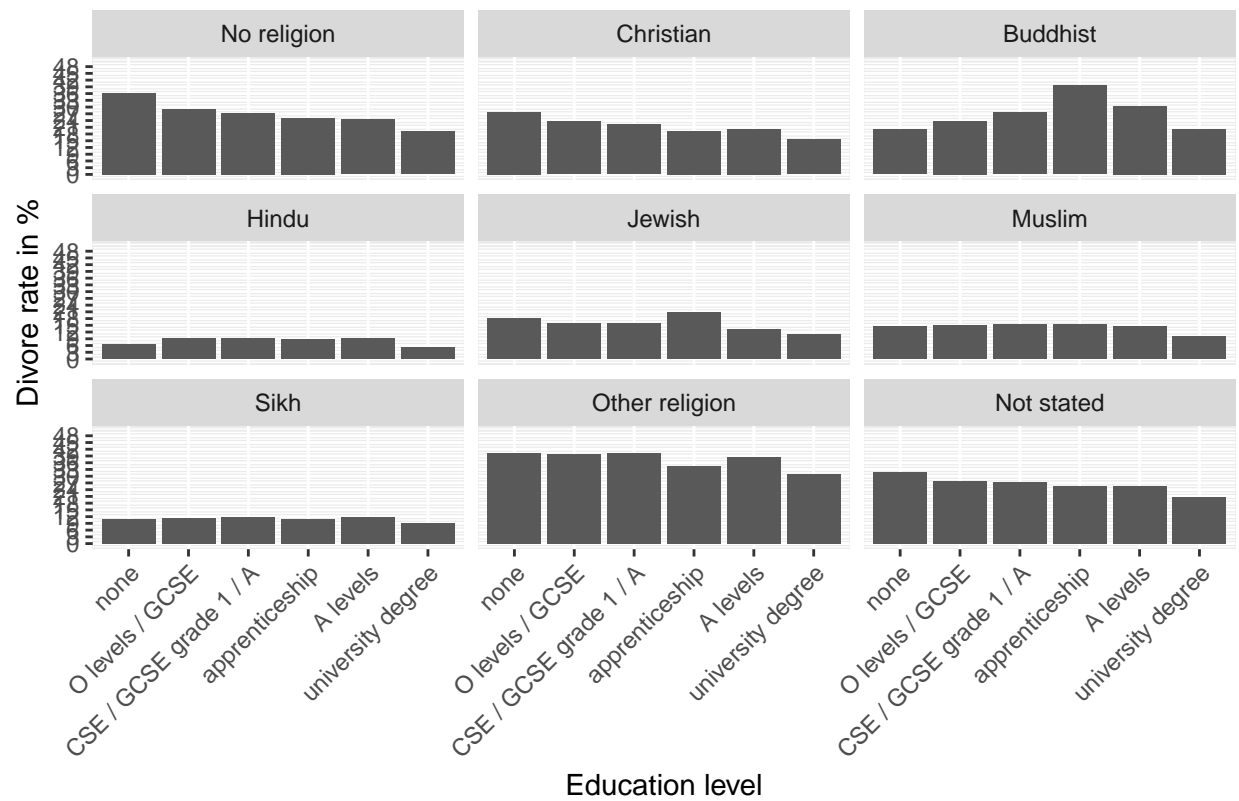
Divorce rate per ethnicity and social grade



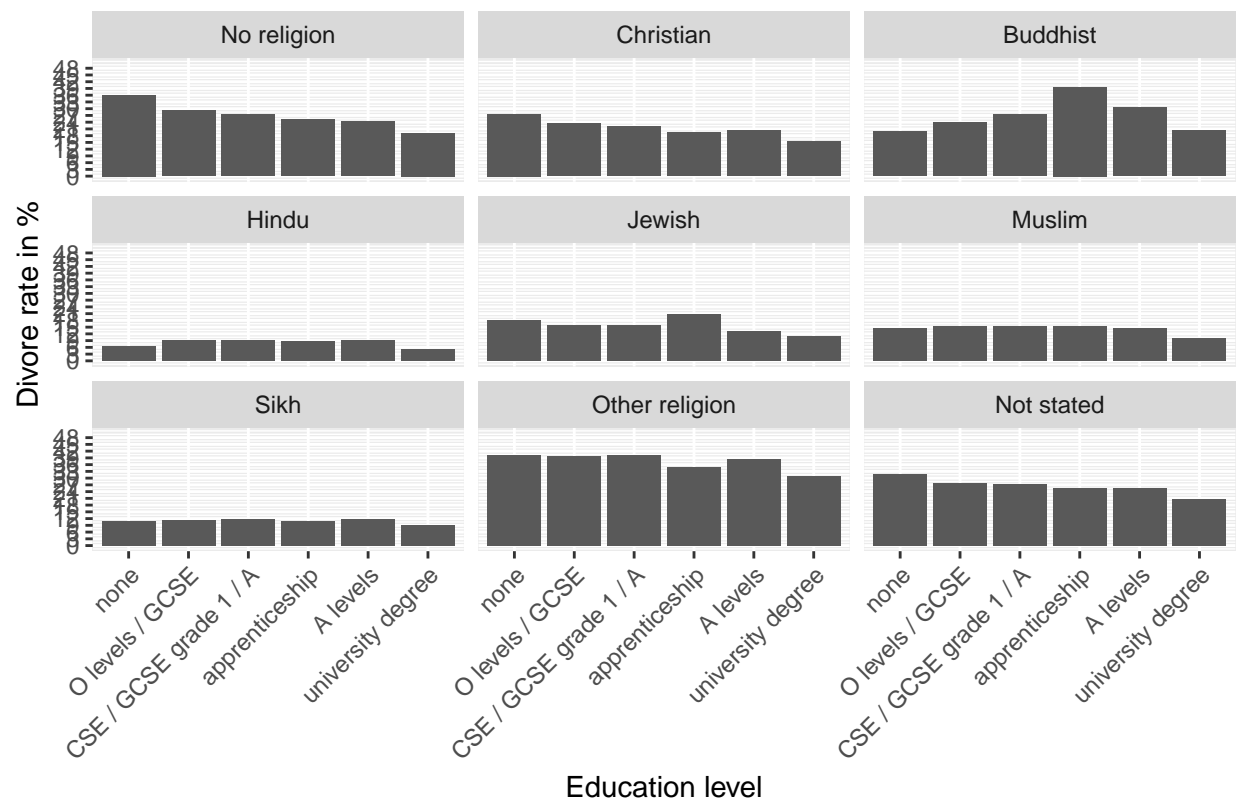
Divorce rate per ethnicity and education level



Divorce rate per religion and education

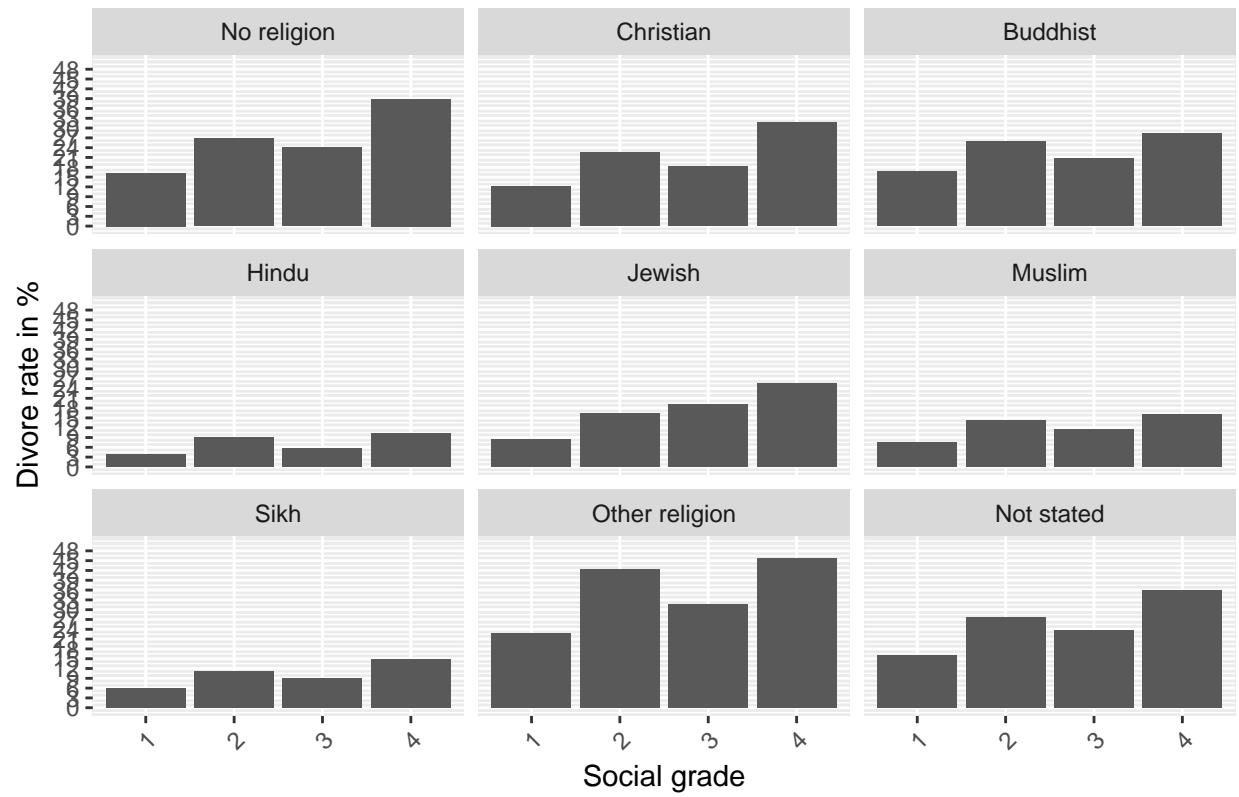


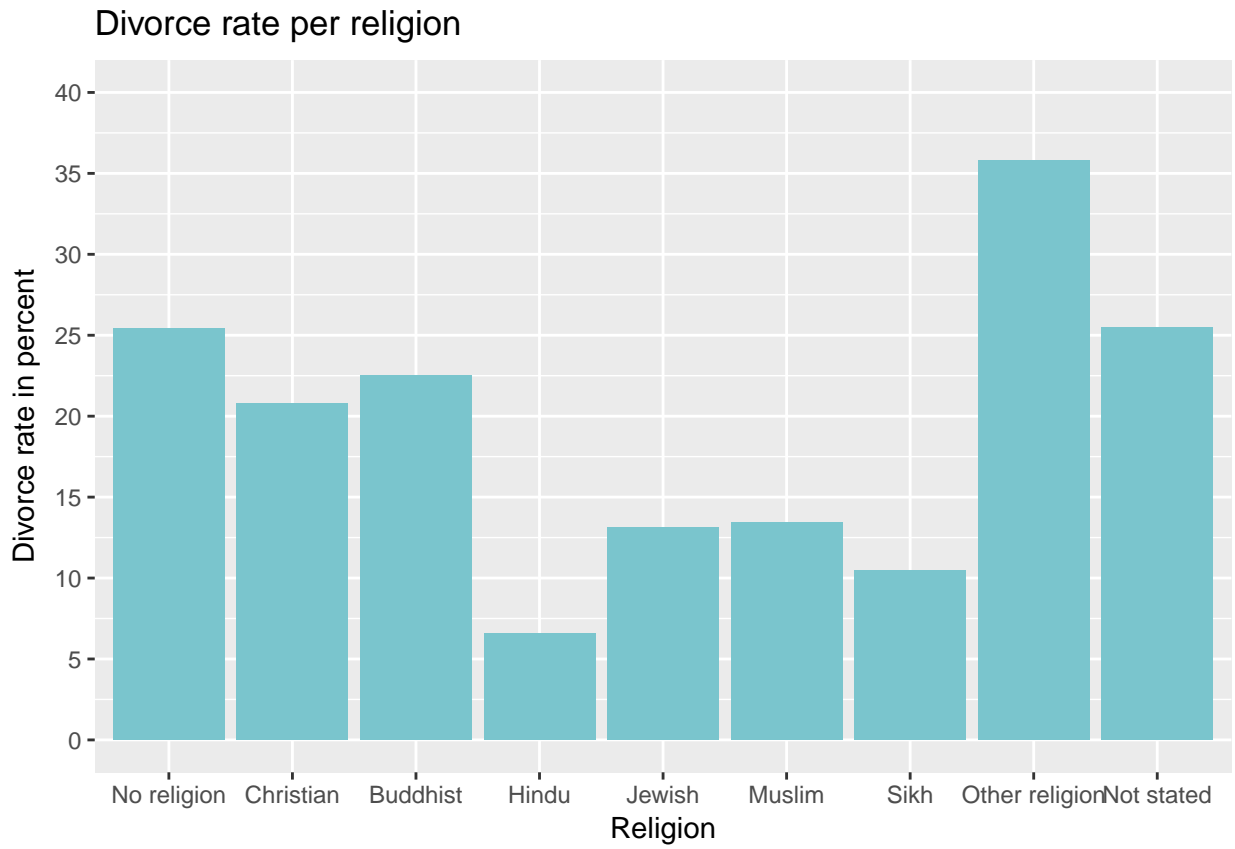
Divorce rate per education level



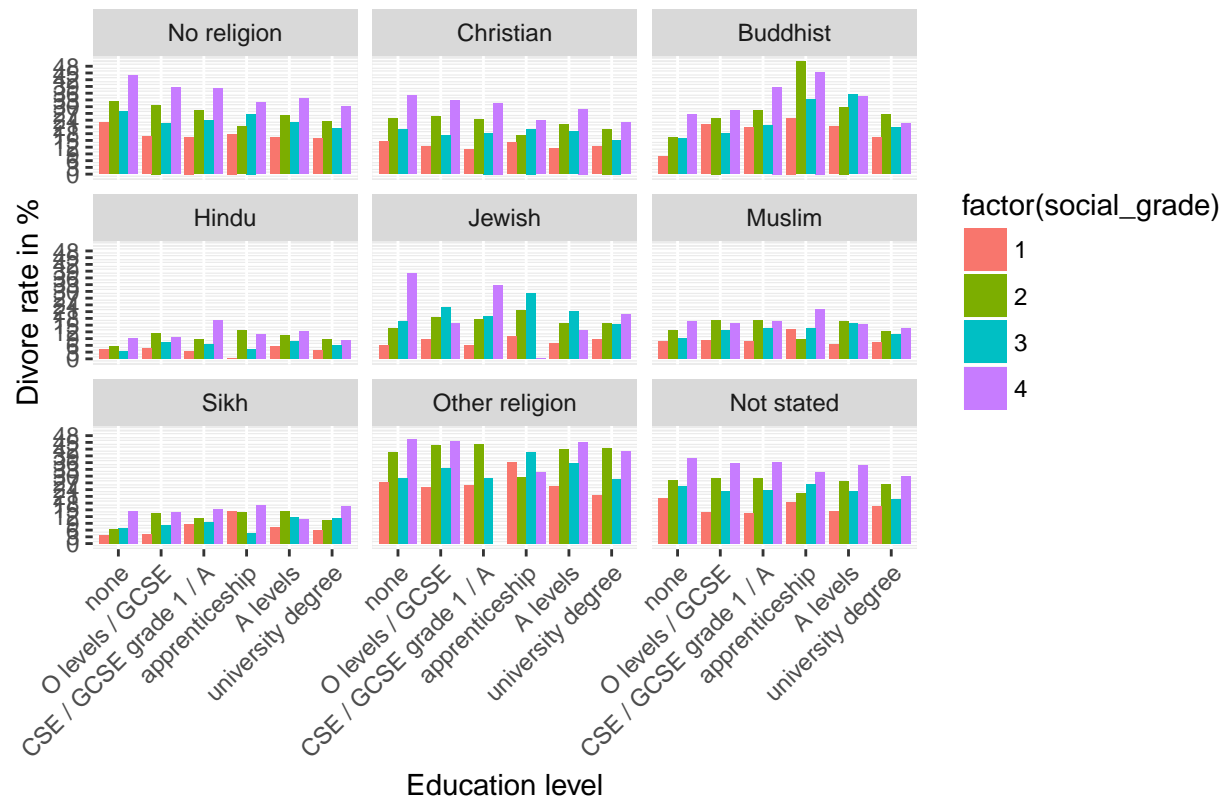
in the different religions the pattern of the divorce rate per education level differe a lot from the overall pattern. Where there is on overall a high divorce rate for people with no education, for the muslims it is way more constant, and for Buddhists it follow a totally different pattern.

Divorce rate per social grade

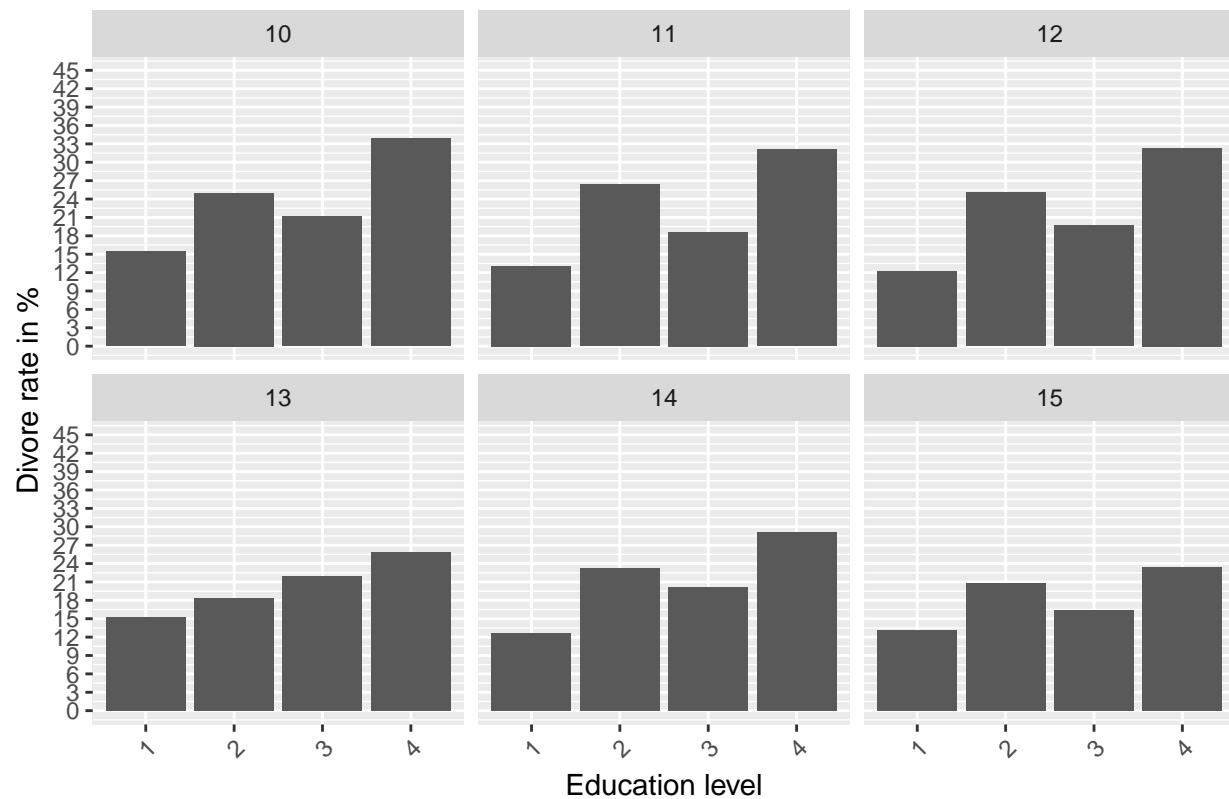




Divorce rate per education level



Divorce rate per education level and region

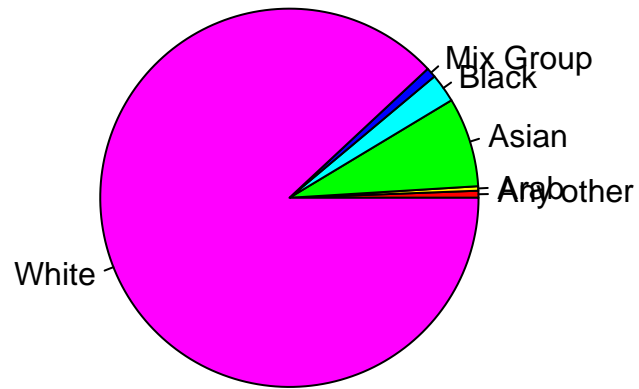


Religion and Culture & Marriage

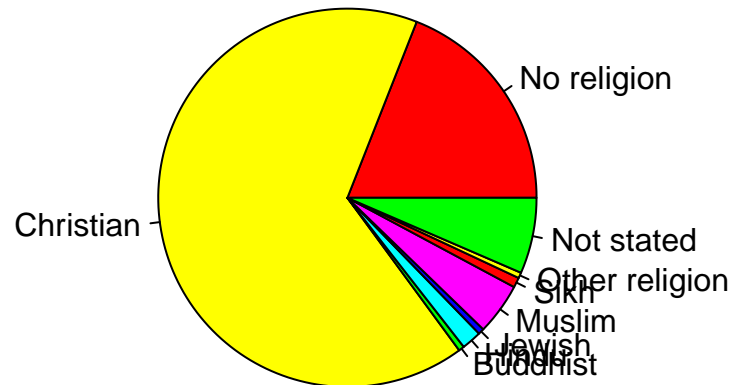
It cannot be denied that the cultural package each person has, is a major factor in his conception of marriage, his expectations toward a relationship and his values about divorce. Thus, in this part, we examine the factors of Ethnicity, Religion and Country of birth to try to determine the effect it has on divorce rate.

These variables are distributed as the following pie charts show :

Pie chart of ethnicities



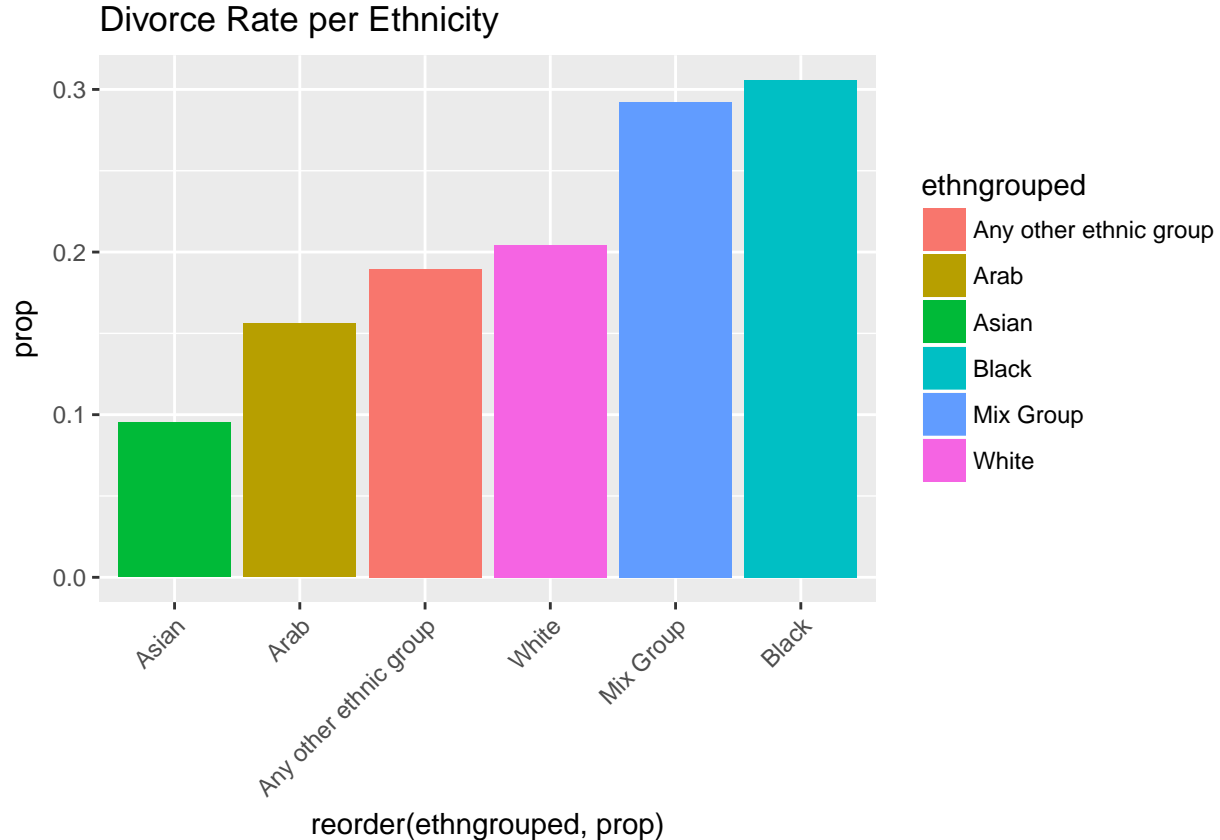
Pie chart of religions



Our assumption is that divorce rate will be higher among white people who are generally considered having more liberal values, and permissive laws for divorce. We also expect that the less committed people are toward religion, the more they are likely to get divorced.

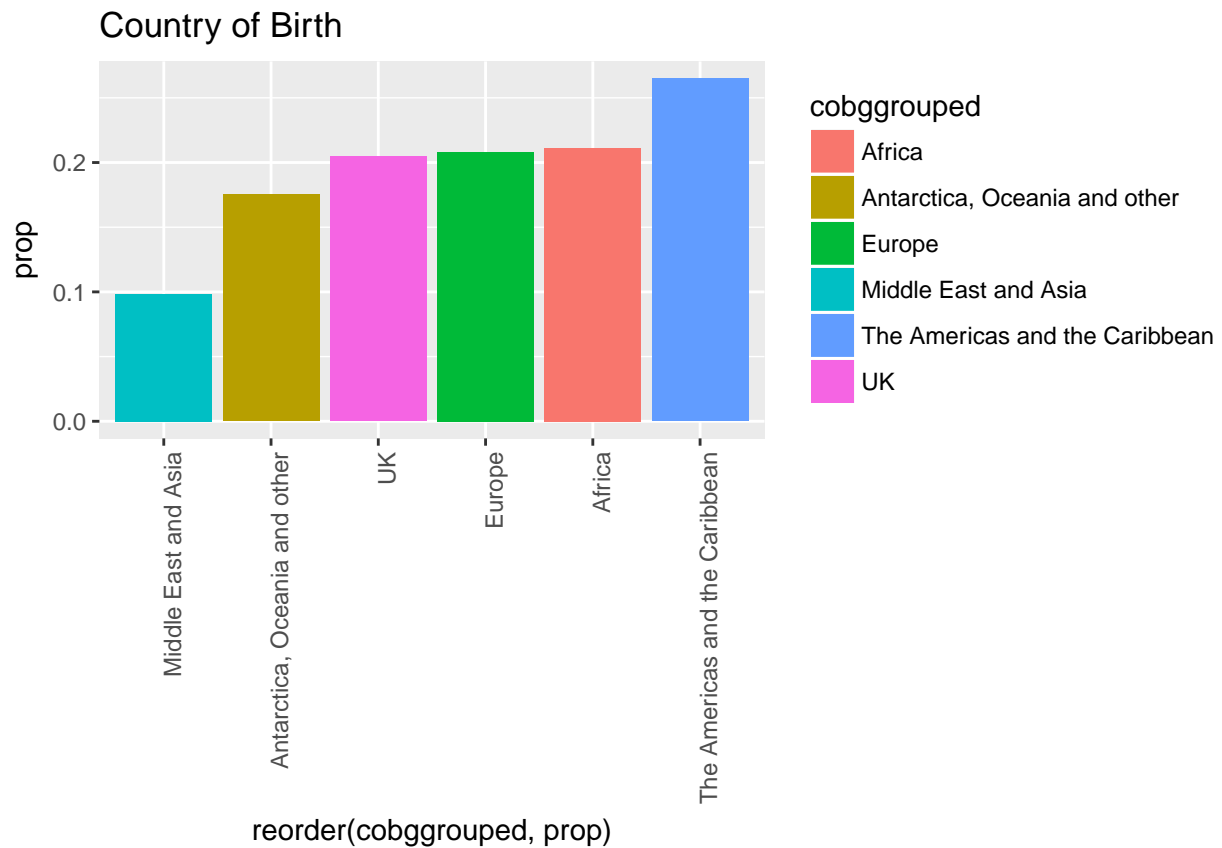
Basic correlation tests between the marital status and the different variables (religion, ethnicities, country of birth) indicate that there is a correlation ($p\text{-value} < 5\%$). Our aim is to have more insight about it.

Ethnicities



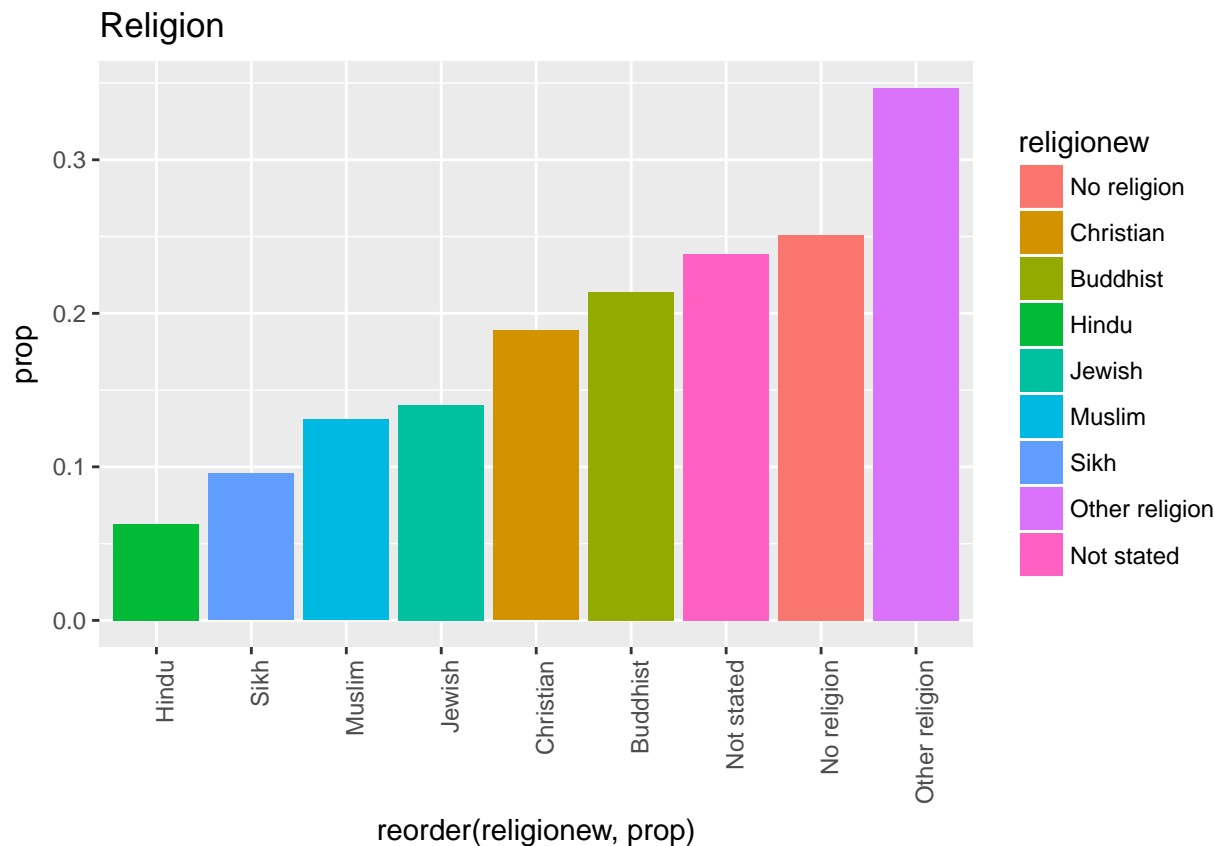
The divorce rate shows us some insight that we expected. For example, Asian divorce rate is significantly lower than for White people. The divorce rate is also very high for a person of mixed ethnicities. This might be explained by the fact that a mixed couple requires a more considerable amount of culture adjustment and understandings, and is more likely to divorce. Parents' value about divorce may thus be transmitted to the child. However, one more unexpected result is that divorce rates are the highest among Black people. If this trend has been the subject of a lot of studies in the US (e.g. <http://www.swaraylawoffice.com/2010/08/divorce-rate-amongst-african-immigrants/>), they are quite scarce about UK. It is commonly known that Black people are still strongly discriminated against in the US, with inequalities in wages, access to education and higher crime rate. But these do not explain entirely the high divorce rate among Black people in UK as we can see in the Education/Ethnicity cross checking, divorce rate is high among Black people no matter what their education level is. Another reason we can think about is the discrepancy between African and European culture. Women in Africa are inferior to men, and confine themselves to domestic roles. In UK, the women have the same rights as men, both socially and economically, giving them more independence and possibly leading to frictions.

Country of birth



The countries of birth tend to confirm the pattern we already found for ethnicities. We can see that the trend of the countries of birth has some similarities with the world divorce ration heatmap you can find following this link: <http://i.imgur.com/dep7XIZ.png>.

Religion



```
## # A tibble: 9 x 7
## # Groups:   religionew [9]
##   religionew `Any other` Arab Asian Black `Mix Group` White
## *      <fctr>      <int> <int> <int> <int>      <int> <int>
## 1   No religion      828   220  6730  1650      2563 241083
## 2    Christian     2206   489 11653 24284      6301 833224
## 3    Buddhist      100     3  4079    38       108  2079
## 4     Hindu       194    12 21902   142        95   277
## 5    Jewish       271    14    60    33         51  6343
## 6    Muslim     2597  3961 41506  4481      1189  4974
## 7     Sikh     1137     6 10006    13         57   167
## 8 Other religion     70    14   1104   136         99  4947
## 9   Not stated     573   312  4627  1997        933 77844
```

Cross-checking the religion and ethnicities variables, we can observe that the religions with the highest divorce rate are 'Other religion', 'No religion' or 'No stated', which might be a sign of more liberal values. Buddhists also have high divorce rate, probably due to its permissive point of view toward divorce (<https://www.budsas.org/ebud/whatbudbeliev/237.htm>). However, the ratio of Buddhists is very low, so their divorce rate has to be taken carefully. Finally, the next religion with the highest divorce rate is Christianity which is not so surprising. Unlike in Catholic Church where marriage is a sacred union, the Church of England has a more open point of view on divorce (which can be related to the historical divorce of Henri VIII). Not to mention that Europeans tend to be less and less committed and involved in religions.

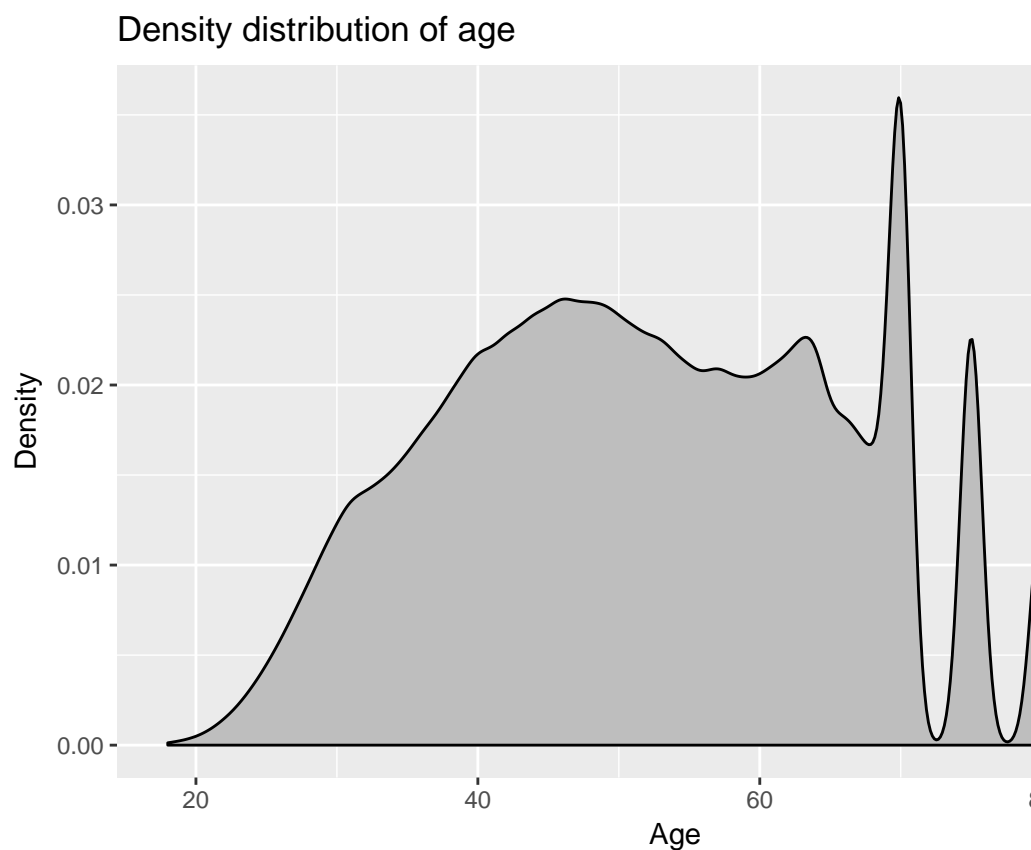
Age & Marriage

Contains information about the age of the observation (person) in years. This variable ranges from 0 to 90 and is treated as a continuous variable. Ages 70, 75, 80, and 85 are clustered into groups each ranging 5 years (i.e. 70 through 74, 75 through 79, etc...). Furthermore, the 90 age group contains ages 90 and above.

For the purposes of this analysis, and to calculate the divorce rate by age, the clustered age groups were equally distributed among the number of ages the group contains. (e.g the 70 through 74 age group was divided into 5 separate age groups 70, 71, 72, 73, and 74 each inheriting the parent group's divorce rate).

Exhibit 1: Density of un-modified age

The chart below, depicts the density distribution of ages in the dataset before any modifications were applied. As expected, instances of the observations' ages are relatively equal among the dataset, which should ensure a



fair representation of all age groups.

Exhibit 2: Divorce rate by age:

The chart below shows the divorce rate (as defined in the introduction) for each of the 83 age groups (18 to 100).

The divorce rate for observations aged 18, 19, and 20 is significantly higher than observations in their 20's. The rationale for this observation is the scarcity (nominally) of married people who are 20 and less, which results in a large relative proportion of divorced people when compared to other age groups. The high divorce rate among young people can also be explained by the often painful and troublesome reality of marriage. <https://www.theguardian.com/lifeandstyle/2014/dec/19/-sp-why-do-so-many-young-marriages-come-to-an-early-end>

The age with the highest divorce rate in the dataset is 48. Likely causes of this statistic are people who got married earlier in their lives and got divorced for reasons stated above, and now belong to this age group, or the shift in societal norms where staying in an un-happy marriage is no longer necessary and

where online dating and other activities are so readily available. <http://www.prince-evans.co.uk/2017/01/divorce-rate-highest-among-aged-40-49/>

This is further supported by the fact that younger people are less likely to be divorced, as those would be early-on in their marriages, while older people have a greater chance of being being remarried and thus will once again be classified as married according to the classifications in the UK census data. Additionally, although divorce rates for people above the age of 53 (Baby Boomers) are rising, this class of society generally has traditional values and belief systems that consider marriage to be the norm. This, therefore, supports the idea that the marginal rate of divorce decreases above the age of 48. <https://www.theguardian.com/lifeandstyle/2013/aug/06/divorces-rise-over-60s>

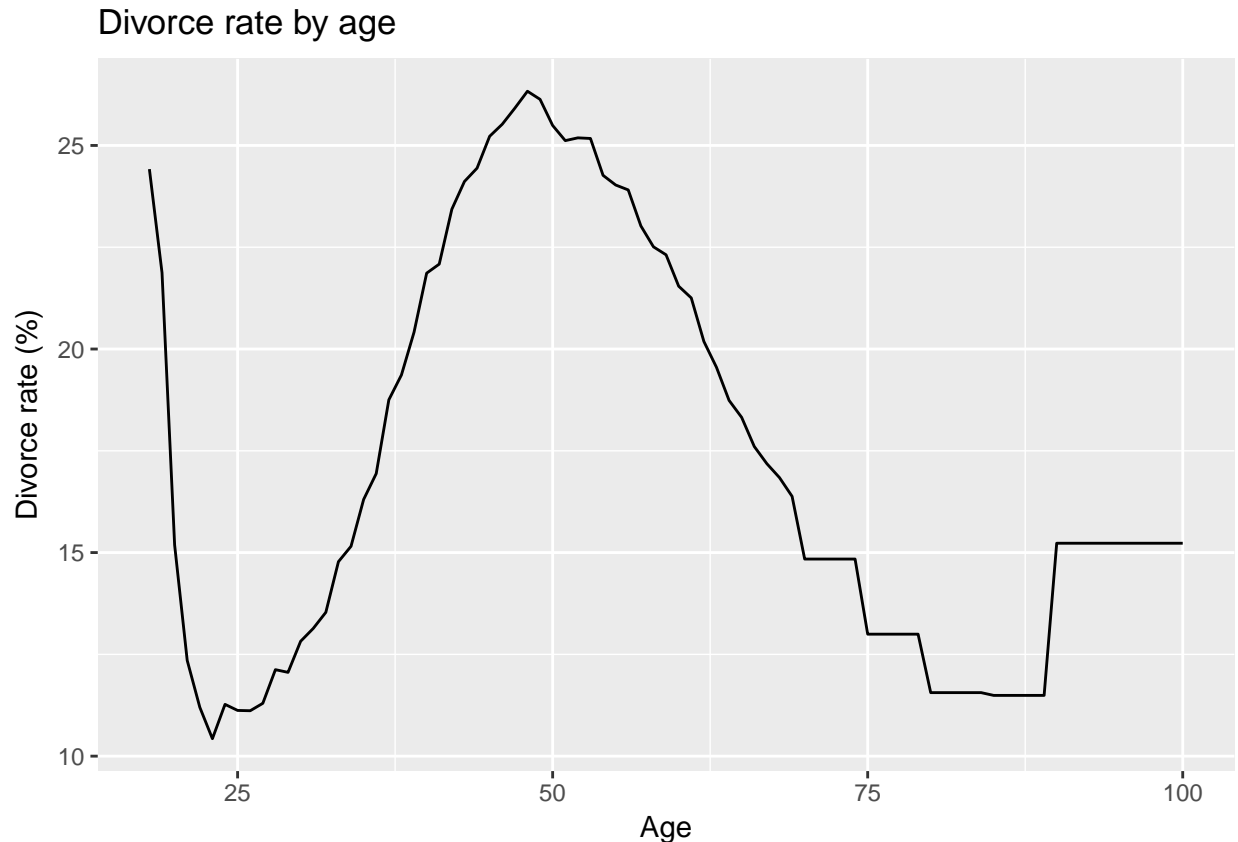
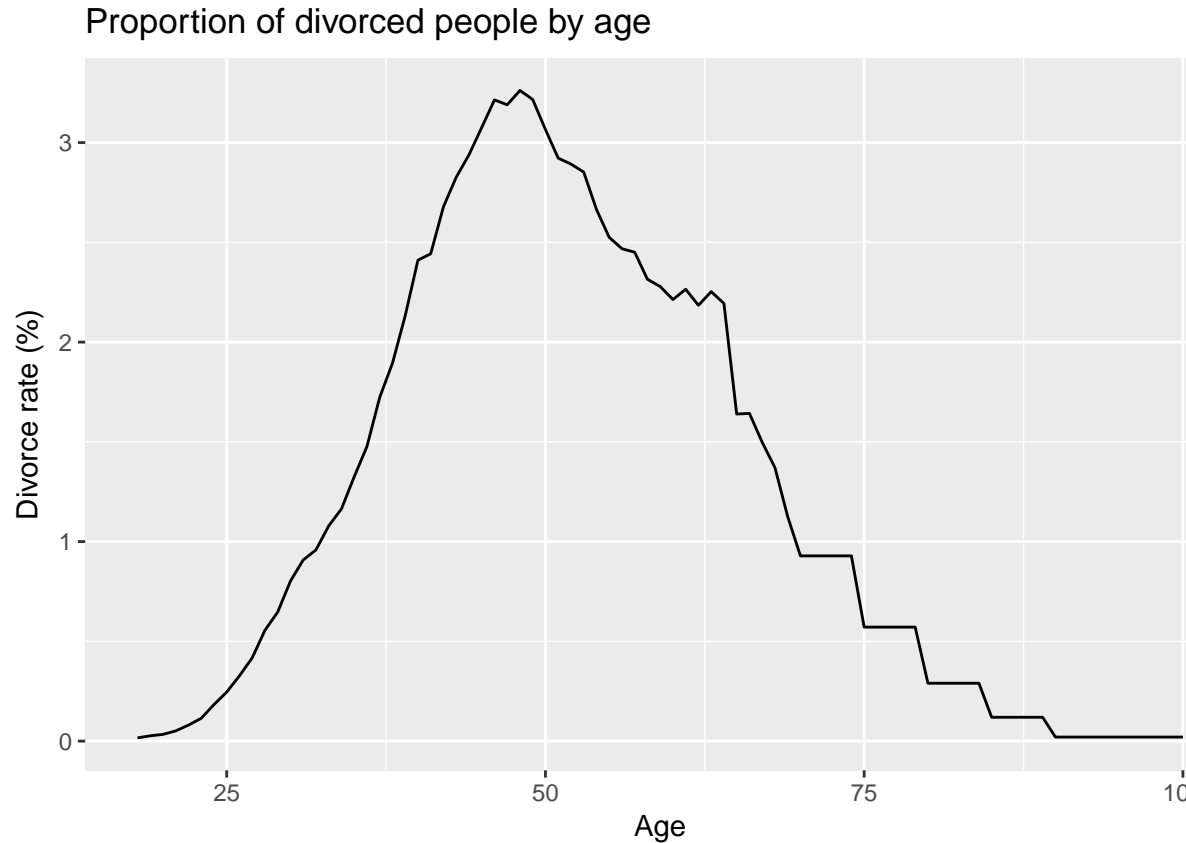


Exhibit 3: Proportion of divorced people by age

The chart above shows the proportion of divorced people for each age group. The chart shows a curve resembling



a normal distribution.

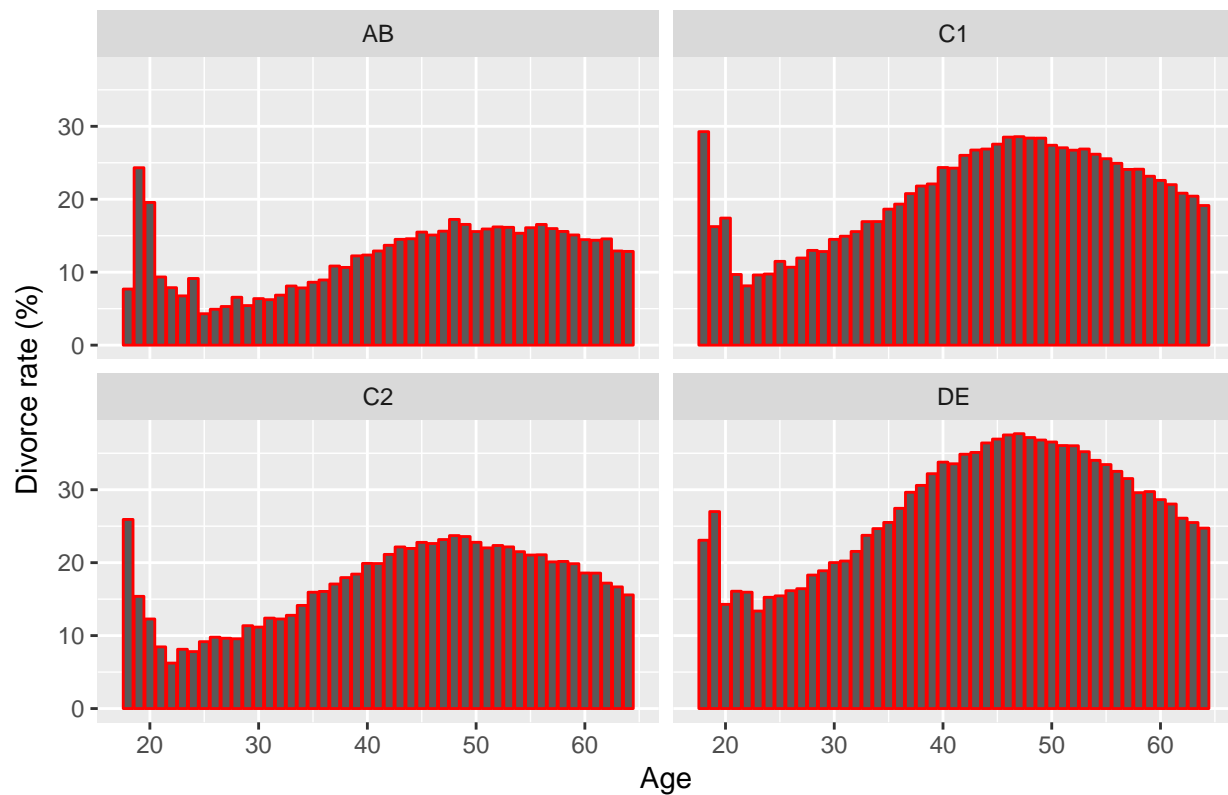
Exhibit 5: Divorce rate by age and social grade

The following chart depicts the distribution of divorce rate by age per level of social class. From the data, it is apparent that the trend observed in the overall sample can also be observed when examining social classes, the divorce rate trends upwards from the lower ages and peaks in the late 40's, this trend then reverses with the divorce rate decreasing with age.

It may be worth noting that while the general trend is consistent accross social classes, one inconsistency does seem to appear in the 18 age group. The people aged 18 in the “AB” social class have a much lower divorce rate of around 8% than the other social classes at more than 20%. This seems to suggest that at young ages, affluence and wealth might potentially provide a cushion for a married couple and alleviate the stresses of financial troubles, which play a big role in divorce.

These findings provide support the notion that divorce rate distribution by age seems to behave similarly regardless of other variables. Further analysis may be required to assess the validity of this claim accross different variables.

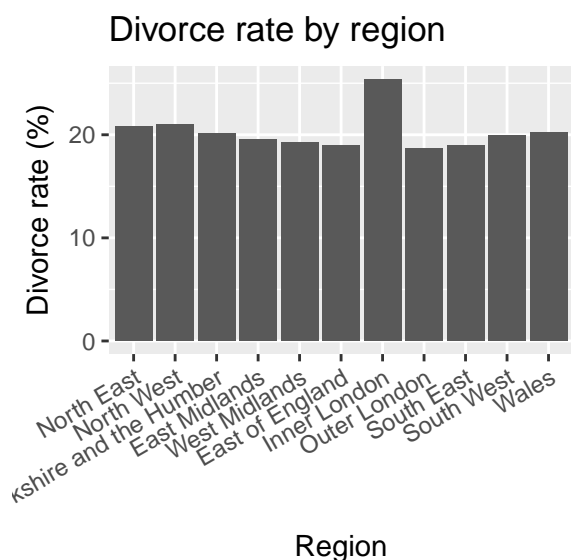
Divorce rate distribution by age per social grade



Region/Geography & Marriage

An article on divorce rate by region piqued our interests (<http://www.dailymail.co.uk/news/article-3201497/Wish-weren-t-ten-divorce-hot-spots-Britain-sea-Blackpool-worst-place-live-want-happy-marriage.html>). It lists the top 10 divorce hot spots and they are all coastal cities. The article states that the factors contributing to the high divorce rate in coastal regions are due to the high deprivation and possibly sheer boredom in winter. In this paper we are keen to explore whether our data attests to the findings of that article.

Our plot shows a significant peak in the percentage of people in bad marriage (either divorced or separated) in Inner London compared to elsewhere. This is the opposite to the aforementioned article. London is anything but a deprived and boring city. The hypothesis testing conducted further confirms that region and marital status are not independent, and particularly Inner London has a higher divorce rate than elsewhere. Inner London has a divorce rate of approximately 25.4%, higher than the average of 19.9% in UK. The difference is statistically significant. Our focus then shifts to explore the factors that are causing such high divorce rates in London.



Could the region people choose to live affect their marriage satisfaction, or is it merely an artifact as a result of other influencing factors? We plot the divorce rates against other variables in our list and wrap them by region. If the variables have no impact, we should observe the same pattern across all regions, with London having a slightly higher divorce rate.

After looking at the relationship between divorce rate and education, social grade, hours worked per week, ethnicity, religion, housing type, number of children in the household, distance travelled to work and means of transport, we found no difference in patterns between Inner London and the rest.

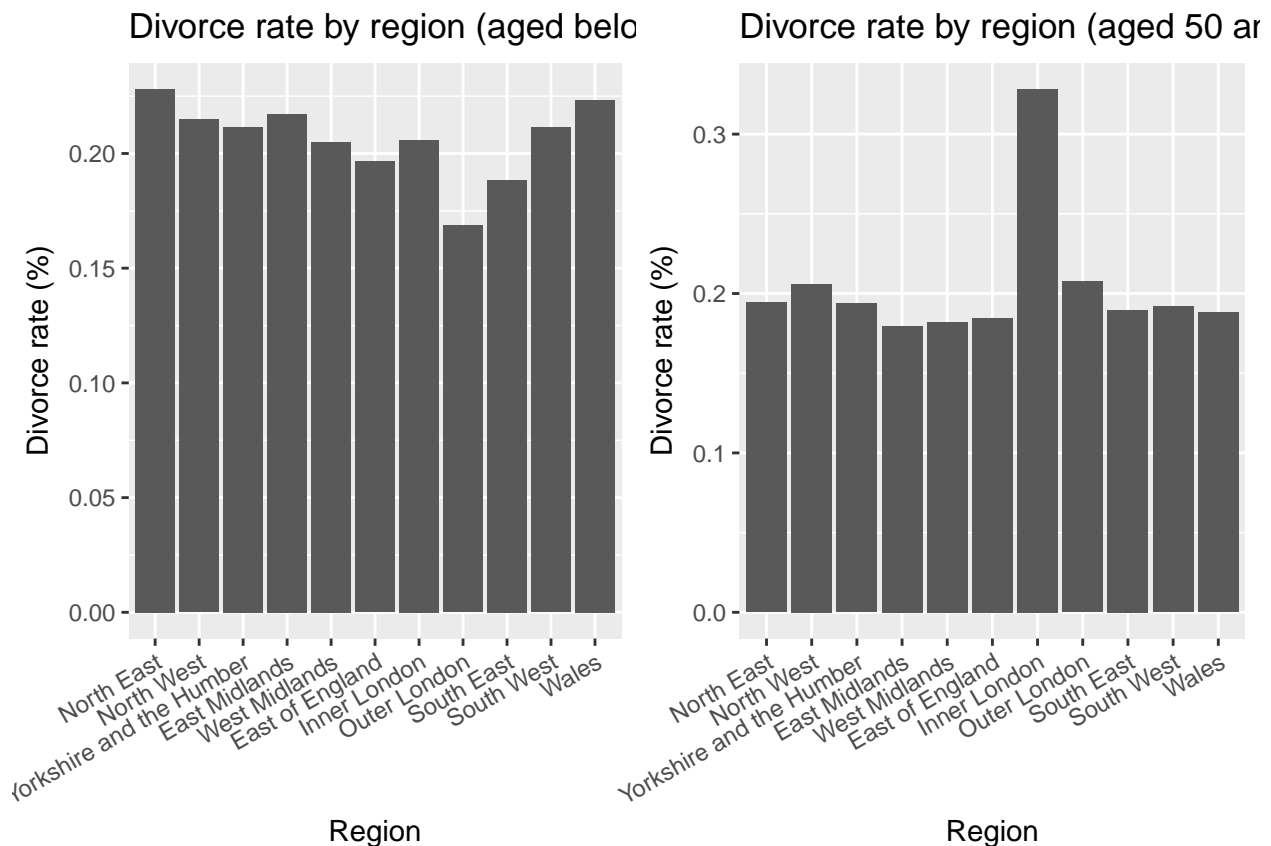
Interestingly we observe a significantly high divorce rate for people aged 50 and above in Inner London. We further break down the population into two groups: age below 50 and age 50 and above. In the first group we no longer see any significant difference in divorce rates across regions, but the difference becomes more apparent in the second group (19.6% for population and 32.8% for aged 50 and above living in Inner London) and is statistically significant.

We have two questions following this observation, firstly why are there so many people getting divorced after 50s and secondly why Inner London?

According to Office for National Statistics in their online publication in July 2017, while the number of divorces drop by 28% from 2005 to 2015, the number of men divorcing aged above 65 went up by 23% and 38% for women. Various factors contribute to the increasing divorce rate among older people (Susan L. Brown and I-Fen Lin 18 Oct 2012). People getting more open-minded and receptive to divorce around them, higher economic independence, and more social interactions with people are among them. More significantly observed in their data is that people in higher order marriages are more likely to end up in divorces than people in first marriage (53.46% compared to 46.54%). And people in their 50s and 60s are also more likely to be in higher order marriages compared to younger people. Unfortunately information on order of marriages is absent in our data.

There is no empirical study on divorce rates in different regions in UK, but we think the reasons mentioned above could all have contributed to the high divorce rates in Inner London. However we must be careful not to confuse the status of being divorced or separated with the instance of getting a divorce. In our data we look at the percentage of people who are divorced or separated at the time of the survey, without information on when and where the divorce occurred. It could also be possible that people in Inner London are less likely to remarry after they are divorced, just like London has the highest percentage of singles aged 30 and above. London is fun and people could be just having a good time being single or in relationships, without rushing into any marriage. We can conclude that there are more divorce and separated people living in London but not infer any causal relationship. Above all, London is the number one Tinder hotspots (thrillist.com, Nov 2015).

Office for National Statistics, July 2017 [<https://visual.ons.gov.uk/marriage-and-divorce-on-the-rise-at-65-and-over/>]
 Susan L. Brown and I-Fen Lin 18 Oct 2012, [<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3478728/>]
<https://www.thrillist.com/travel/nation/where-tinder-passport-users-are-swiping-the-most>



Dependent children

When people get married they are often in love and happy, most married couples look forward to parenthood and having children could even strengthen their marital bond. However empirical study shows that the effect of children on marital stability could be very complex. Children increase the stability of marriage only when they are preschool-aged (Becker et al. 1977), as they grow older, their binding force in a marriage weakens; the greater number of children, the smaller possibility of divorce, but the relationship is inversely quadratic rather than linear suggesting the stabilizing effect is the strongest with the first and second child (Xu, Yu and Qiu 2015).

The birth of an infant often brings couples back together who would have otherwise divorced. Experiencing divorce often place young children at risk for a variety of problems, which may persist into adulthood (Morrisox and Coiro, 1999). Therefore married couples may choose to stay in marriages for the wellbeing of their children but when children are old enough they decide to part their ways.

We are keen to explore if we observe the same pattern in our data (1) as the age of youngest child grows, the more likely the couple will get divorced/separated (2) the more children a couple have, the less likely they are to be divorced/separated.

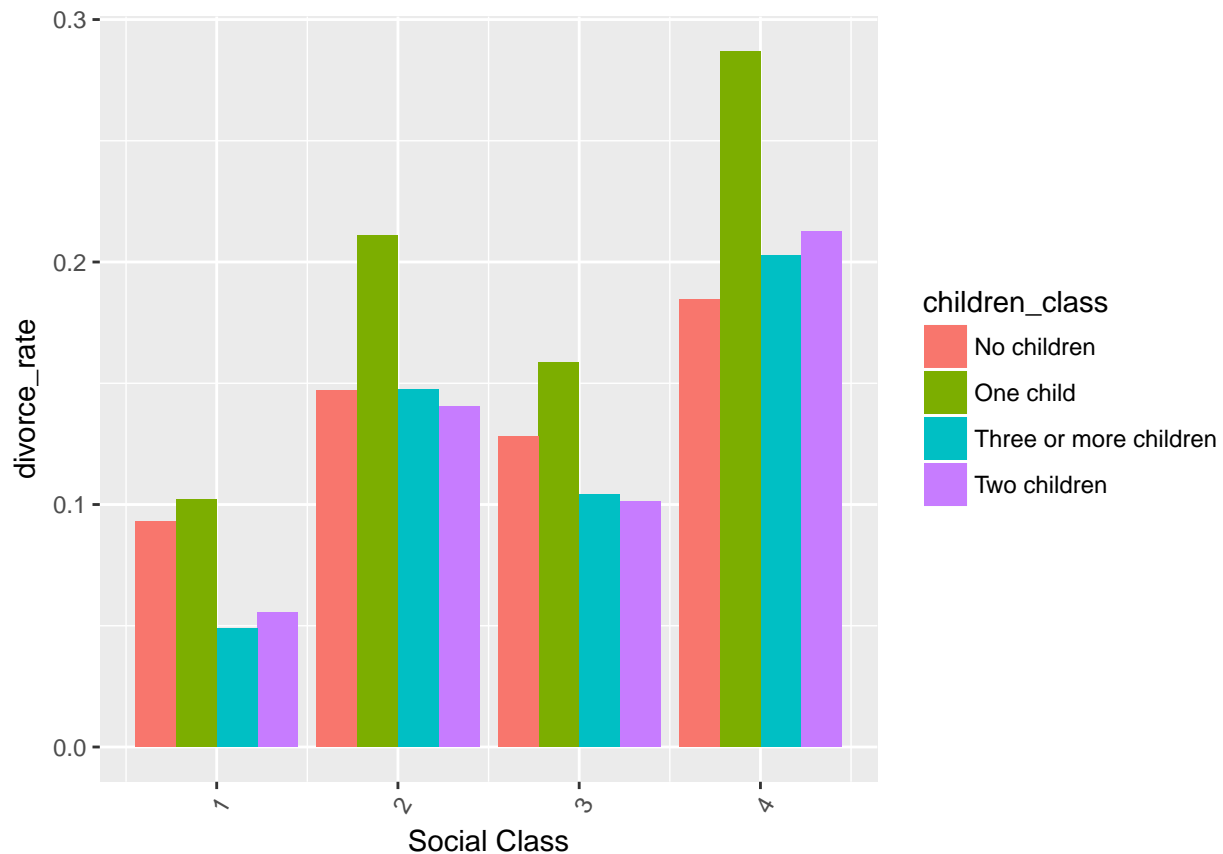
Becker, Gary S., Elizabeth Landes, and Robert T. Michael. 1977. "An Economic Analysis of Marital Instability." *Journal of Political Economy* 85:1141-87.

Waite, Linda J, and Lee A Lillard. 1991. Children and marital disruption. *American Journal of Sociology* 96(4):930–953.

Qi Xu^{1*}, Jianning Yu² and Zeqi Qiu³. 2015, The impact of children on divorce risk. Xu et al. *The Journal of Chinese Sociology* (2015) 2:1

Donna Ruane Morrisox and Mary Jo Coiro, Parental Conflict and Marital Disruption: Do Children Benefit When High-Conflict Marriages Are Dissolved?. *Journal of Marriage and the Family* 61 (August 1999): 626-637

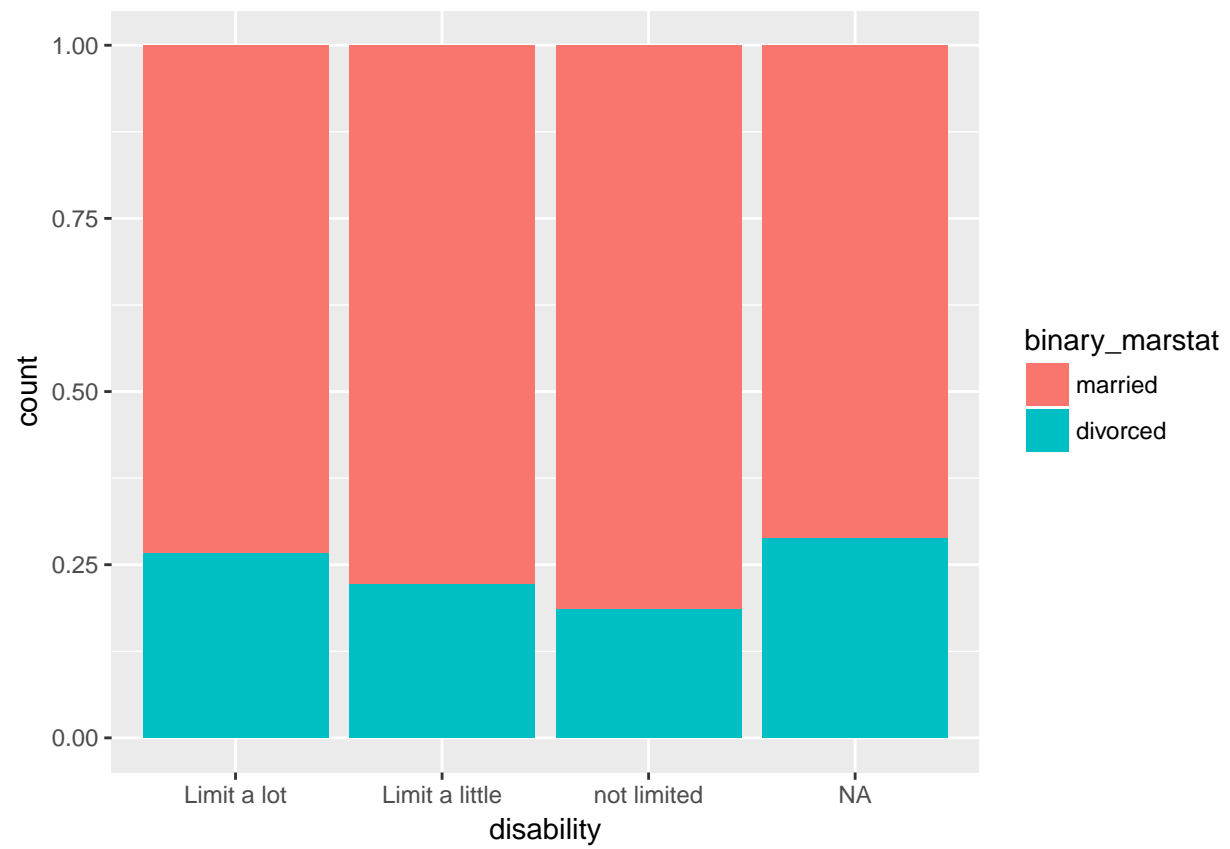
Dependent Children and Social Grade

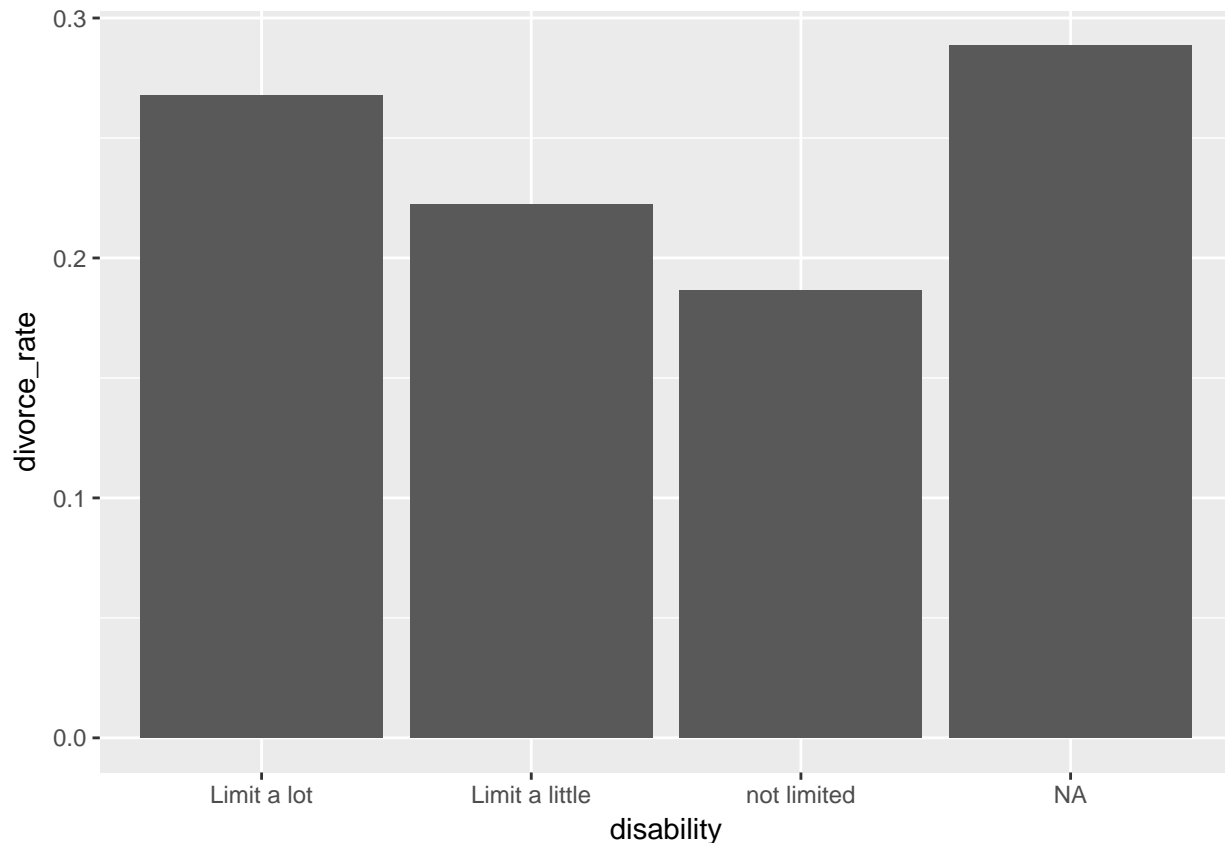


Health & Marriage

```
##
## Pearson's Chi-squared test
##
## data: mine$binary_marstat and mine$disability
## X-squared = 5428.6, df = 2, p-value < 2.2e-16
```

A chi-square test of the two variables indicate that the correlation between marital status and long-term health problem is significant. This finding suggests that the two variables are not independent of each other. Long-term health problem will influence marital status.





The graph shows that the less severe the long-term health problem, the less the divorced rate.

Logistic Regression

A comprehensive regression analysis of all correlated variables

Method As the dependant variable in question is a dichotomous variable (divorced or not divorced), we chose to build a logistic regression model instead of a linear regression model. Then we selected the variables discussed above as dependent variables. The variables presented thus far in this report are the variables hypothesised to have the greatest effect on a person's likelihood of being divorced in the sample. Our goal was to use the variables we examined to predict whether a person will be divorced.

Overall, we experimented with a total of 8 variables to find the best Logistic Regression Model: 1. Social Grade 2. Education 3. Age 4. Religion 5. Health 6. Number of Children 7. Ethnicity 8. Whether the person lives in London

The best fit was determined by adding and removing the variables above until the best combination of the aforementioned metrics was obtained.

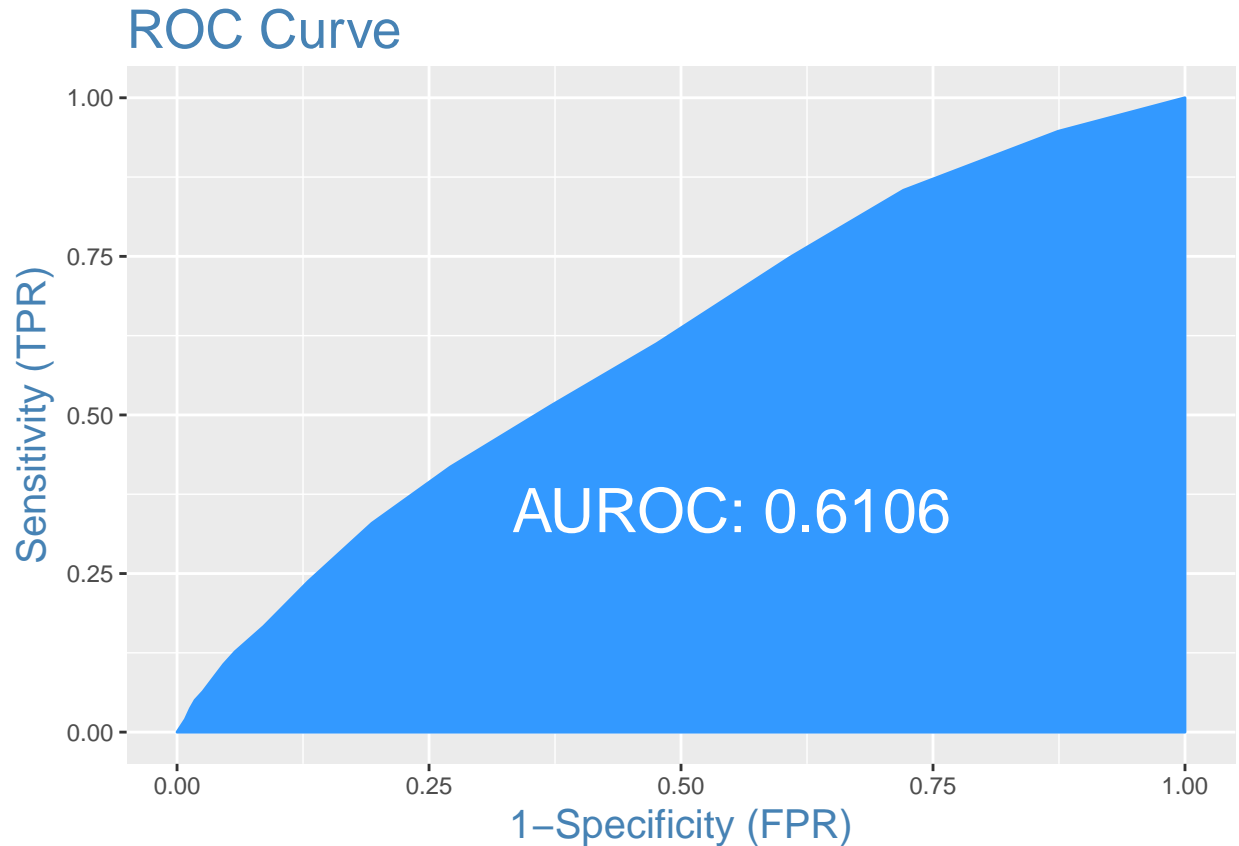
As a measure of the overall model's explanatory power, two main measures were use: 1) c-statistic or AUC (area under ROC curve) 2) The accuracy of the predictions above a certain threshold (the mean divorce rate for our purposes)

In addition, we also examined the multicollinearity of our model to ensure its validity. In order to understand how good our model is, we partitioned the original dataset into 70% training data and 30% test data. After we used the training data to train the model, we then used the test data to compare our models.

Based on our analysis, the best model contains 4 variables: Social Grade, Education, Health, and Age.

ROC Curve

A close look on the ROC Curve can give us more insights. Shown by the sensitivity and 1-specificity index, the increase of the True positive Rate of our model is slightly faster than the increase of the false positive rate. Since the greater the AUC (Area Under the Curve) the better the predictive ability of the model, the AUC of our model 0.6105 indicates that our model is fair but not perfect.



Accuracy

Next, we calculate the model's accuracy by comparing it to the mean of the test dataset. The result indicates that the accuracy rate is 58.89%. The accuracy rate tells us how accurate our predicted result compared to the actual result is. Our model will get about 59 correct results within 100 examples.

```
## [1] 0.5889018
```

Multicollinearity

Finally, we examined the multicollinearity in the model by using the Variance Inflation Factor. According to the result below, all indicators are very close to 1 and well below 5. Therefore, all predictors within this particular logistic regression are not linearly related. We can safely conclude that the multicollinearity is very low and the model is valid.

```
## social_grade    education    disability      age
##      1.286545      1.324527      1.087492      1.067298
```


Limitation

Unfortunately, we recognize that our model does not outperform the base prediction because the mean divorce rate is about 21%. This means that if we predict all people are not divorced, we will yield a higher prediction result. We suspect that this result is a problem rooted in the original massive census dataset because there are a total of 121 variables, and we are likely to miss some potential drivers of the marital status.

Conclusion

Nevertheless, our finding is still significant because we identified a few strong drivers to unhappy marriages. Among all variables we discussed in this report, Social Grade, Education, Age, and Health are the most important factors for predicting if a person is likely to be divorced.

Hypothesis Testing and Prediction Model (Helene's)

Conclusion

This paper has examined variables hypothesised to affect divorce rates in the UK. According to the analysis performed above, it is evident that divorce rates have multiple determinant factors, which themselves are potentially co-dependant. This analysis does not attempt to single out a specific variable or group of variables as the main determinant of divorce rate, but examines potential affecting variables.

Lacking any information about a particular person, age would be a good first predictor of marital status. It has been shown that people in their 40's have the highest divorce rates in society. This is true regardless of social status.

It is evident that social class and education level are amongst the biggest influencing factors of divorce rate as depicted by the strong inverse relationship of these variables with divorce rates. This trend holds true across all ethnicities and religions, which further supports the strength of the variables.

References

Main Github Repository: <https://github.com/dlouhasha/TheGreatWork>

Appendix

A list of all variables and descriptions