
Data Exploration Project

Name: Nabilah Anuwar

Table of Contents

INTRODUCTION	2
Questions	2
DATA WRANGLING	2
DATA CHECKING	4
DATA EXPLORATION	4
How has marriage (divorce) rate changed over time?	4
What are noticeable trends related to marriage in Australia?	6
How does marriage rate differ by location (States and Territory) and time?	9
How does marriage rate differ by location (Australia and Other Countries) and time?	10
CONCLUSION	10
REFLECTION	10
BIBLIOGRAPHY	11
APPENDIX	12
Appendix 1: Example data viewed in Excel before wrangling	12
Appendix 2: Example step 1 result of Data Wrangling	13
Appendix 3: 2018 File I - viewed using head() function in RStudio	13
Appendix 4: Line graph for each countries marriage rate per 1000 people	13
Appendix 6: Change in rate of change from 2008 to 2017	13
Appendix 7: Average marriage rate per 1000	14
Appendix 8: Average Marriage Occurrence by Day of the Year - Annotated	16

INTRODUCTION

Over the years, people seem to believe that there are an increasing number of divorces and younger people refusing to get married. However, same sex marriages have been legalized in some countries which make us believe that the number of marriages should increase. More people getting education also seems to affect the number of marriages as they put career over marriage. Change of ideas in the 20th century where divorces are understandable also affect how people perceive divorces. These events make us believe that divorce is supposedly to increase and marriage, though not as much, decrease over the past few years.

Questions

1. How has marriage (divorce) rate changed over time?
2. What are noticeable trends related to marriage in Australia? E.g.: marriage celebrant, marital status, age group, ethnicity.
3. How does marriage (divorce) rate differ by location (Australia and other countries) and time?

DATA WRANGLING

1. Divide and arrange data in excel
 - a. We look at the data in Excel and see that there are many sections within one sheet, multiple merged cells, and combination of heading with a description (Example: "Total(a)(b)" we will turn it to just "Total"). If we put it inside RStudio or Tableau we will have a very messy output. So, we rearrange the data using Excel.
 - b. Since there are many information presented in the excel sheet, we will divide it to multiple tables.
 - c. We format the cells according to its data type
 - i. Numbers of counts will be presented as integers with no decimal places
 - ii. Rates is presented by 2 decimal places
 - d. When finished, we will export it as csv and continue wrangling in RStudio.
 - e. To see the example data before and after wrangling please refer to the appendix.
 - f. We separate each excel file to:
 - i. Number of marriages registered and its types of marriage (first marriage or remarriage; by minister of religion or civil servants; country of birth of couple) each year
 - ii. Number of males and females first marriage/remarriage each year
 - iii. Marriage occurrence by day
 - a) The total column and row from the original dataset are removed in this step as it will ease validation process
 - b) Reason of removal is that total is subject to deviation which may not equal to its total sum of the relevant cells
 - iv. Number of marriage and its types of marriage based on its state and territory
 - v. Number of divorces granted with children from divorce and country of birth of couple per year
 - vi. Marriage rate per 1000 people per country for 2008 to 2017
2. In RStudio we will combine the tables from the 2018, 2014, and 2010 tables to get statistics from 2018 to 2008 to compare trends over the past 10 years as each table have information for the starting year and 4 years of the previous years.
 - a. Using tidyverse library we import the csv files
 - b. Using tidyverse select function, we will combine the 2018, 2014, and 2010 tables selecting only years within 2018 – 2008. Table example is available in appendix 3.
 - c. Then using merge function, we will combine the tables
 - d. 2018 and 2014 table have information of 2014 statistics. We will choose the newer version which is in 2018 database. This will also be applied for 2010.
3. We want the years to be in a single column, so we need to reverse them using spread and gather function

Initially it looks like this:

Description	2008	2009	2010	2011	2012	2013	2014	2015	2016
born in different countries	35629	36185	37925	38488	38686	37553	38408	36240	37374
both born in australia	70521	69643	68586	67882	69046	65605	66074	61565	64552
both born in same overseas country	12460	14210	14588	15319	15439	15728	16188	15762	16439
civil celebrants	77142	80340	83799	85311	88599	86301	89891	85124	90451
first marriage both partners	82847	84780	86328	86863	88540	86076	87811	81715	85559
first marriage one partner	19969	20107	20129	20415	20089	19339	19439	18542	18913
ministers of religion	41545	39740	37254	36375	34612	32601	31247	28409	27893
remarriage both partners	15940	15231	14715	14476	14604	13549	13950	13339	13928
total marriage	118756	120118	121173	121754	123243	118959	121197	113595	118401

Figure 1 No of Marriage Occurrence table

Then we apply the functions

year	born in different countries	both born in australia	both born in same overseas country	civil celebrants	first marriage both partners	first marriage one partner	ministers of religion	remarriage both partners	total marriage
2008	35629	70521	12460	77142	82847	19969	41545	15940	118756
2009	36185	69643	14210	80340	84780	20107	39740	15231	120118
2010	37925	68586	14588	83799	86328	20129	37254	14715	121173
2011	38488	67882	15319	85311	86863	20415	36375	14476	121754
2012	38686	69046	15439	88599	88540	20089	34612	14604	123243
2013	37553	65605	15728	86301	86076	19339	32601	13549	118959
2014	38408	66074	16188	89891	87811	19439	31247	13950	121197
2015	36240	61565	15762	85124	81715	18542	28409	13339	113595
2016	37374	64552	16439	90451	85559	18913	27893	13928	118401
2017	36208	60336	16388	88061	81511	17967	24880	13475	112954
2018	38062	64353	16724	95003	85954	19584	24178	13645	119188

Figure 2 Final No of Marriage Occurrence Table

4. Datasets for the information about number of marriages based on states and territory, it is separated in 11 datasets for each year. This is what the data looks initially

Description	NSW	VIC	QLD	SA	WA	TAS	NT	ACT	Total
total marriage	40,008	25,620	23,504	7,122	12,766	2,356	827	1,385	113,595
first marriage both partners	29,256	18,775	16,346	5,047	9,117	1,611	544	1,029	81,715
first marriage one partner	6,427	4,115	3,995	1,122	2,098	412	165	207	18,539
remarriage both partners	4,320	2,731	3,167	956	1,555	345	113	153	13,342
ministers of religion	11,607	6,514	4,797	1,798	2,700	479	142	375	28,419
civil celebrants	28,398	19,059	18,708	5,327	10,060	1,878	692	1,002	85,115
both born in australia	21,097	13,115	13,999	4,517	5,830	1,836	408	766	61,565
both born in same overseas country	6,518	4,490	2,124	660	1,679	59	66	168	15,762
born in different countries	12,394	8,010	7,380	1,949	5,238	461	351	456	36,240

Figure 3 Excel Table of States and Territory

- a. We will remove the total column as the Australian Bureau of Statistics (2020) mentions that the total may not present the total of the cells as it is subject to a deviation. We will not use that and instead use the counts that is already presented.
- b. Before combining we will make the dataset into a long form using tidyverse gather function.
- c. Adding a column for its year
- d. Then combining all to one dataset
- e. As before, we then use spread function

Description	state	count	year
born in different countries	ACT	513	2018
born in different countries	ACT	515	2017
born in different countries	ACT	393	2016
born in different countries	ACT	456	2015

Figure 4 States and Territory after gather function

state	year	born in different countries	both born in australia	both born in same overseas country	civil celebrants	first marriage both partners	first marriage one partner	ministers of religion	remarriage both partners	total marriage
ACT	2008	494	1051	129	1024	1151	304	650	219	1674
ACT	2009	458	998	106	978	1090	263	584	209	1562
ACT	2010	468	979	148	1027	1168	242	567	185	1595
ACT	2011	485	885	141	997	1092	255	507	164	1511
ACT	2012	500	926	176	1039	1172	233	558	197	1602
ACT	2013	466	850	144	981	1080	212	478	168	1460
ACT	2014	484	876	159	1113	1102	241	408	180	1523
ACT	2015	456	766	168	1002	1029	207	375	153	1385
ACT	2016	393	625	160	909	860	202	264	115	1175
ACT	2017	515	944	214	1282	1212	285	391	181	1674
ACT	2018	513	873	181	1303	1096	271	262	194	1562
NSW	2008	11879	21612	5583	23613	27999	6429	15470	4709	39137

Figure 5 States and Territory after spread function

to make its description as its dataset columns.

5. Marriage occurrence by day
 - a. With the totals removed, we need to convert it to a long form first using gather function
 - b. We then add a new column for the year it is set in
 - c. Combining all the 11 datasets then we continue to data checking
6. After merging we export the files as csv from RStudio. However, this is done after we check the data.

DATA CHECKING

1. First, we check that the tables have the correct column format using the str function. From here we can see that there is a mistake as the year column should be as numeric. So, we use the transform function to the dataset.

```
'data.frame': 11 obs. of 10 variables:
 $ year          : chr  "2008" "2009" "2010" "2011" ...
 $ born.in.different.countries : num 35629 36185 37925 38488 38686 ...
 $ both.born.in.australia      : num 70521 69643 68586 67882 69046 ...
 $ both.born.in.same.overseas.country: num 12460 14210 14588 15319 15439 ...
 $ civil.ccelebrants          : num 77142 80340 83799 85311 88599 ...
 $ first.marriage.both.partners : num 82847 84780 86328 86863 88540 ...
 $ first.marriage.one.partner   : num 19969 20107 20129 20415 20089 ...
 $ ministers.of.religion        : num 41545 39740 37254 36375 34612 ...
 $ remarriage.both.partners     : num 15940 15231 14715 14476 14604 ...
 $ total.marriage               : num 118756 120118 121173 121754 123243 ...
```

Figure 6 str function detail

2. To ensure that there are no NAs we use the drop_na function to drop any rows that have a missing value

```
'data.frame': 11 obs. of 10 variables:
 $ year          : num 2008 2009 2010 2011 2012 ...
 $ born.in.different.countries : num 35629 36185 37925 38488 38686 ...
 $ both.born.in.australia      : num 70521 69643 68586 67882 69046 ...
 $ both.born.in.same.overseas.country: num 12460 14210 14588 15319 15439 ...
 $ civil.ccelebrants          : num 77142 80340 83799 85311 88599 ...
 $ first.marriage.both.partners : num 82847 84780 86328 86863 88540 ...
 $ first.marriage.one.partner   : num 19969 20107 20129 20415 20089 ...
 $ ministers.of.religion        : num 41545 39740 37254 36375 34612 ...
 $ remarriage.both.partners     : num 15940 15231 14715 14476 14604 ...
 $ total.marriage               : num 118756 120118 121173 121754 123243 ...
```

Figure 7 after column correction

- a. In Marriage rate per 1000 people per country for 2008 to 2017, we drop any countries that has NA as its values which meant it is done before the gather function.
 - b. In the number of males and females first marriage/remarriage each year we substituted "-" with 0 as it indicates that there are no given events of that category that time
 - c. Marriage occurrence by day file has all month include 31 days, to tackle that we substitute 'wrong days' with empty cells at the beginning to result in NA when inputted in RStudio. When removed we have 4,018 observations which is correct.
3. When using Tableau, despite having all columns correctly formatted in RStudio and Excel, they still may import it as characters. Double checking whenever new dataset is inserted and changing the column type accordingly.
4. Whenever an import is done in RStudio, checking the number of variables and observation it is supposed to have and adjusting it accordingly
 - a. A variable had an empty column after the gather function. Using the tidyverse select function we select only the necessary columns

DATA EXPLORATION

How has marriage (divorce) rate changed over time?

Using Tableau, we try to see how marriage/divorce change over time based on their number of occurrences.

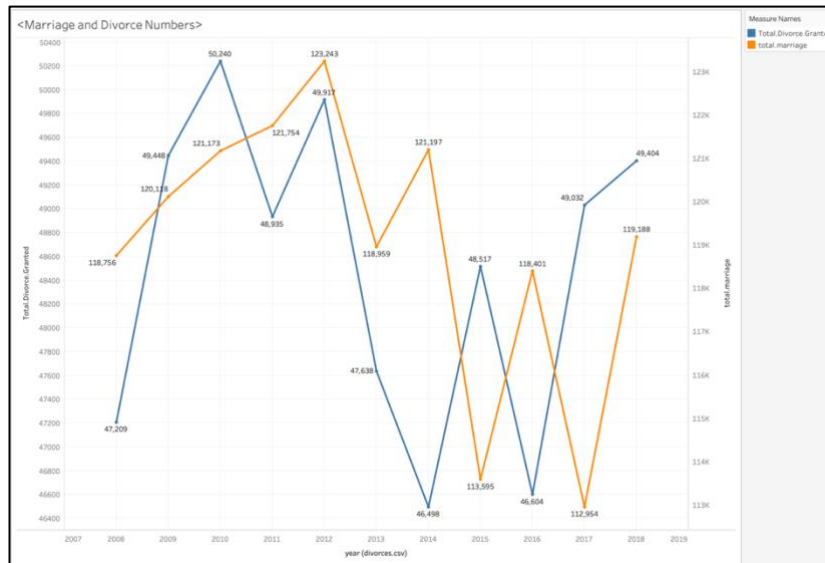


Figure 8 Marriage and Divorce Numbers

occurrence over the years. We see that the number of divorces has been fluctuating over the years even previous to 2012, different to marriages. It is noticeable that the number increase and decrease consecutively. However, there is a constant movement of decrease/increase where it is the same as the previous year (decrease was followed by a decrease, and otherwise). These years are 2010, 2014, and 2018. It is interesting that every four years it follows the behaviour of the previous year. It can mostly be coincidental however, further study in human behaviour and social trends would help uncover the factor that correlates to it.

It is noticeable from this graph is that the number of divorces is almost half of the number of marriages registered. Though there seem to be no correlation to the numbers of divorce and marriages except some instances on where it negatively correlate and other positively correlate. Positive correlation seems to happen in the years 2008 to 2013 as one mimic the other's movement. Negative correlation happens for a couple of times in the years 2014 to 2018. We further test this by using RStudio pearson correlation test (STHDA, n.d.).

```

>>>{r}
cor.test(d$total.divorce.granted, m$total.marriage, 'two.sided', 'pearson')
>>>

Pearson's product-moment correlation

data:  d$total.divorce.granted and m$total.marriage
t = 0.43262, df = 9, p-value = 0.6755
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4999496  0.6840385
sample estimates:
cor
0.1427317

```

Figure 10 Correlation of Marriage and Divorce Numbers

year	divorce_rate	marriage_rate
<dbl>	<dbl>	<dbl>
2009	4.74	1.15
2010	1.60	0.88
2011	-2.60	0.48

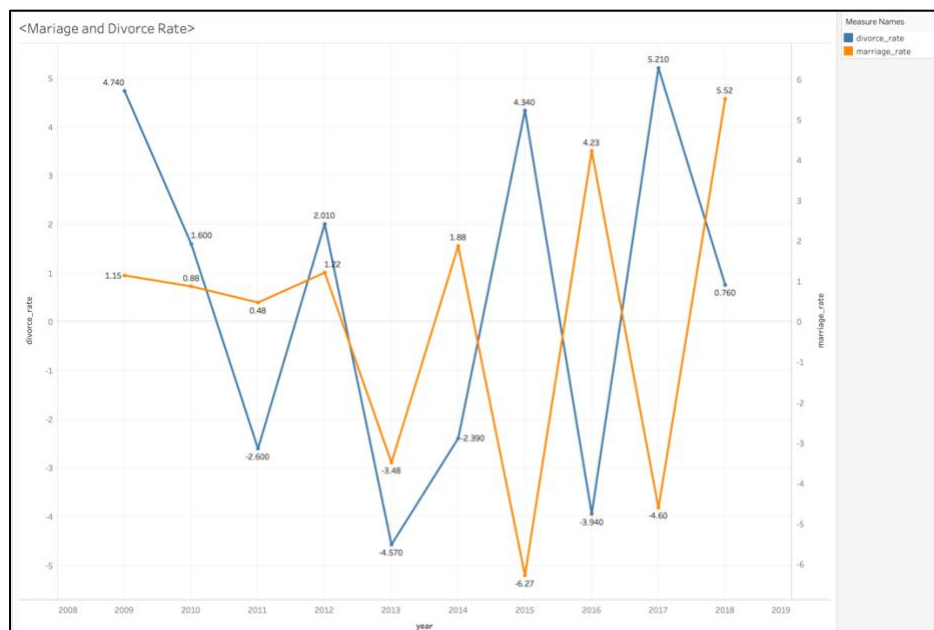
Figure 9 Marriage and Divorce Rate per year

As seen here the number of marriages was steadily increasing from 2008 to 2012. Then, it plummets in 2013 almost to the same count in 2008. It increases halfway, almost the same as 2010 count. It then dropped significantly in 2015 and increase halfway again in 2016 only for it to plummet lower in 2017 and increase again in 2018. It is very interesting how in the first four years the rates have been consistent yet from 2013 onwards the rate fluctuates.

Continuing, we then look at the number of divorce

The test shows that there is no significant correlation between the two as their p-value of 0.6755 is not below 0.05 significant value. Thus, no correlation.

Using RStudio, we calculate the growth rate for marriage and divorce. Using dplyr, tidiverse, and method by Bob (2018) we get a rate of change. In this result the year presented is the rate of change for the year before till the year present (example 2018 present rate of change for 2017 to 2018).



Next is we try to see the connection between the rate of marriage and divorces. First, lets look at the marriage rates. Though its count was increasing, their growth rate was decreasing from 2009 to 2011. Their rate has been fluctuating since 2008. In the end of 2018, their rate is higher than ever before. In the beginning, divorce rate seems to decrease yet in 2014

onwards it fluctuates more extreme than marriage rates. 2018 divorce rate only seems to lower the rate to almost zero rather than an extreme negative rate different to its previous year. Seeing both together, again we see some kind of correlation as they have some similar movement.

However, when we test it using pearson correlation we do not get a p-value within the 0.05 significant level. If there is a significant value, their correlation would be negative as their correlation coefficient is -0.3576.

```

cor.test(r$divorce_rate, r$marriage_rate, 'two.sided', 'pearson')

Pearson's product-moment correlation

data: r$divorce_rate and r$marriage_rate
t = -1.0832, df = 8, p-value = 0.3103
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.8058105 0.3510391
sample estimates:
cor
-0.3576323

```

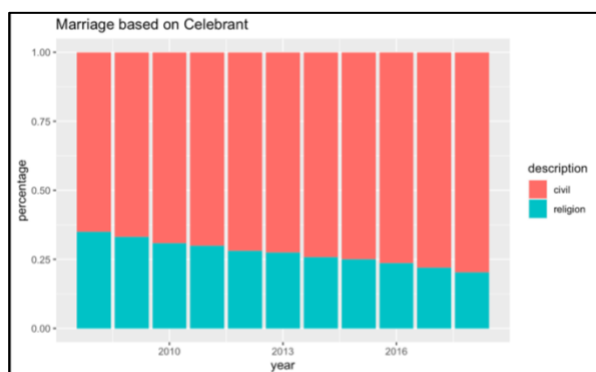
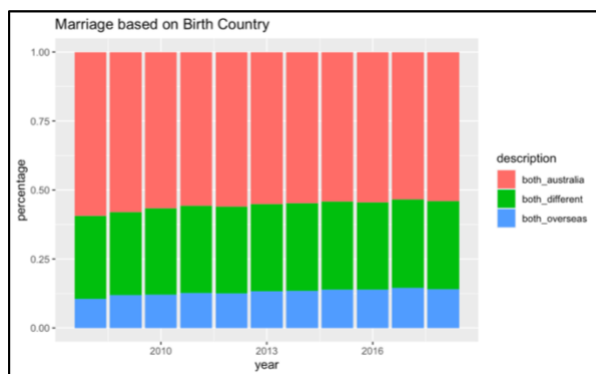
Figure 11 Correlation between Marriage and Divorce Rate

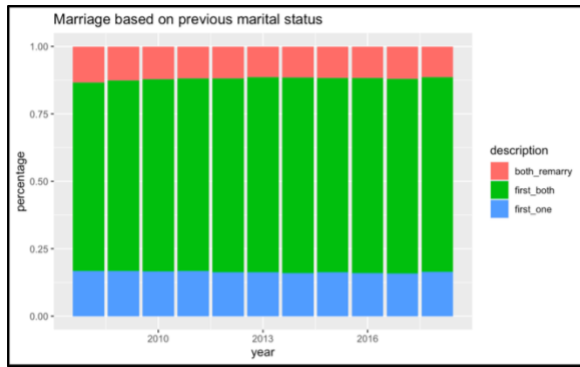
What are noticeable trends related to marriage in Australia?

Using RStudio we analyze the trend of marriage based on birth country, celebrant, previous marital status, and age group using ggplot2 bar charts.

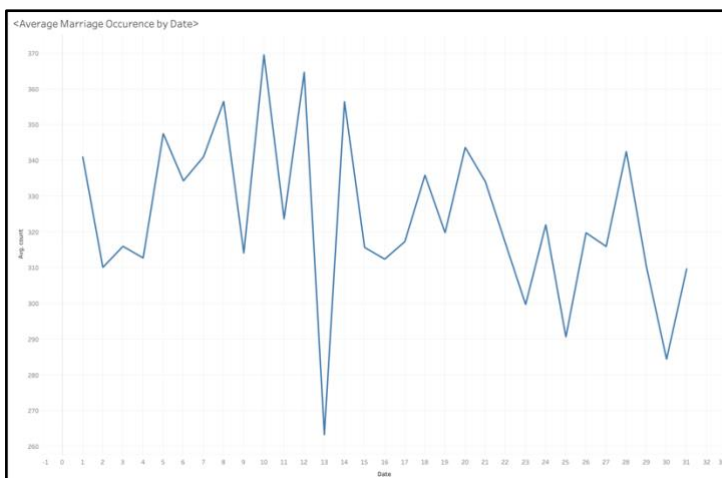
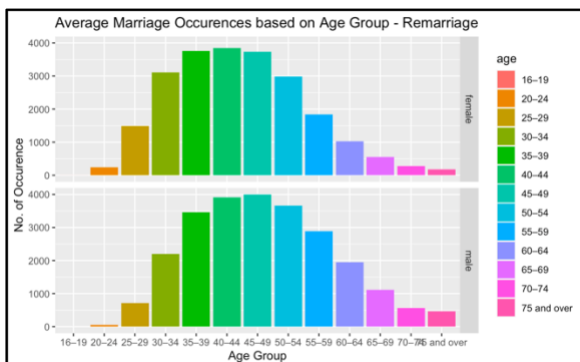
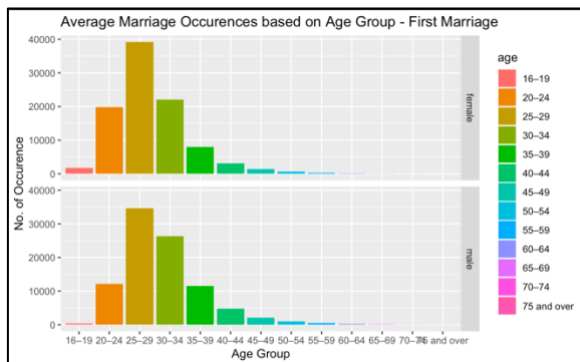
Since 2008, marriage between Australians is what approximately 50% of the marriage occurred. This is expected as our target location is indeed Australia. Marriage from different birth countries has increased over time as well as those who are both from overseas countries. This could mark that the number of visitors that settled from another country has increased overtime.

In this modern era people are more open minded to each other's beliefs. This tendency may be the cause of why the number of marriages that is have a ministry of religion as its celebrant decreased over time. Easy Wedding (n.d.) has mentioned in their article that civil celebrants wedding is much simpler. People would prefer





again or the number of people marrying once is just higher.



simplicity over complexity; thus, they might be more interested in having a civil wedding. It can also be possible that the number of marriages with people of different faith affecting the increase of civil marriage.

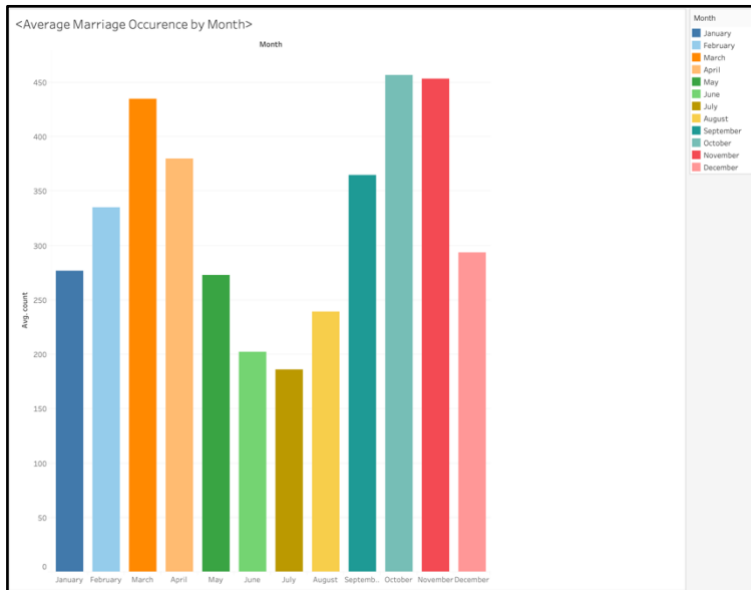
There seems to be no significant changes in previous marital status in marriage. Both partners that marry for their first time dominate throughout the years. Followed by one of the partners already married to have a slightly larger amount than if both partners remarry. This may mean that people who marry once may not have the tendency to marry

again or the number of people marrying once is just higher. Female and male have the tendency to marry at their late twenties. Though for males, it is not as high as females as their number is reaching to 40,000 average. Female aged 16 to 24 are more likely to have their first marriage than that of male. However, male in their thirties are more likely to have their first marriage than that of the women. This can be affected by the belief that women should marry before the age of 30.

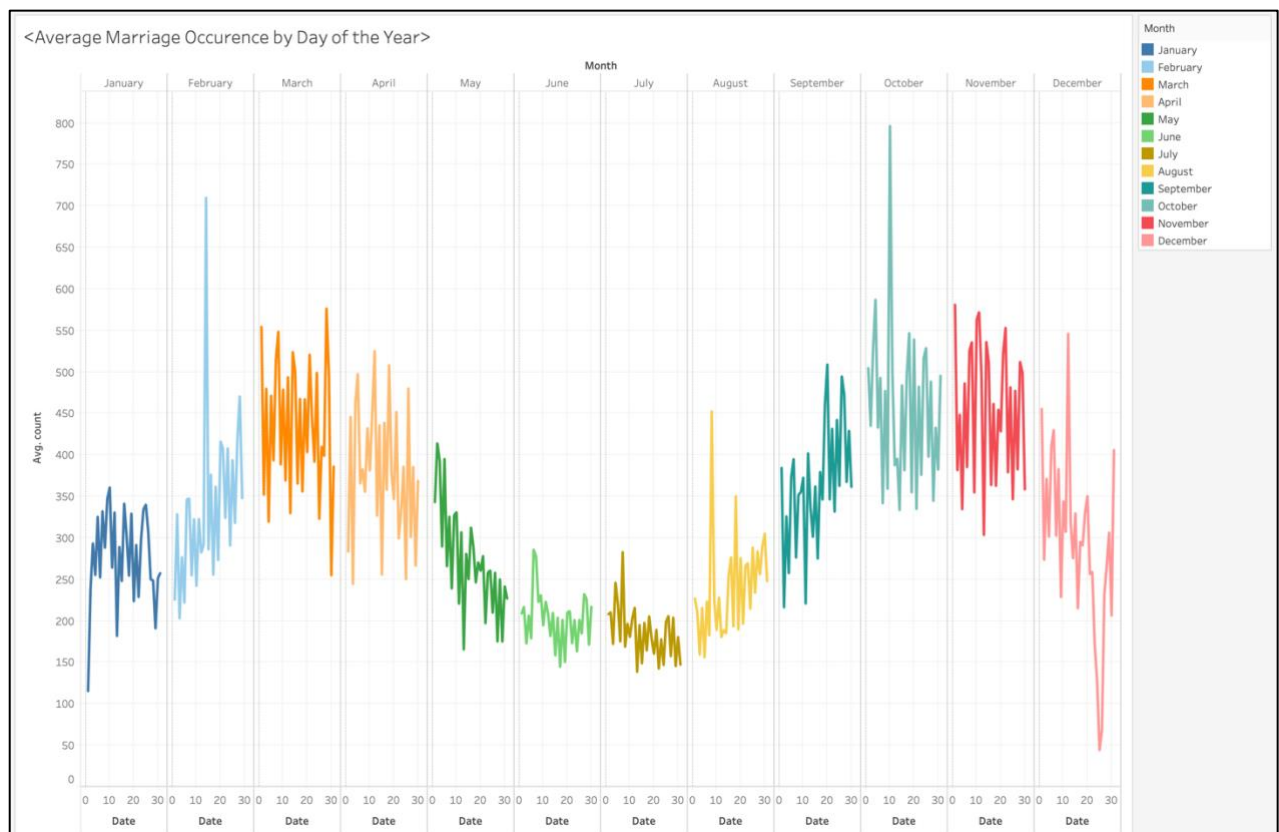
Females in their early forties have the highest tendency to remarry, whilst men are in their late forties. The age gap is not much of a difference for men and women. The number of males remarry after the age of 50 is significantly higher than that of the female. The number of those that are in 60 to over age group have around twice the size of female in the same age group. The numbers for male in their fifties are also significantly higher than of female.

From both graph we can see that males have the tendency to have a marriage at a later age than females.

Marriage is a sacred thing and getting the perfect date is very important to some people. Made with Tableau, the graph of average marriage occurred by date, we can see that there aren't many people who would like their wedding to be in the 13th. This may be caused due to superstitious beliefs that the number 13 is unlucky. We can see that the date 10th is the most preferable and second to it is 12th. Second least preferable is the 30th where it is near end of the month.



Looking at the average marriage occurrence by month we can see that there is some kind of 'peak season' and 'low season'. October seems to be the highest, followed by November and March. The low season are found in between May to August, with July as its lowest month. Weather may affect this decision as December to February is summer, too hot, and May to August is winter, too cold, in Australia. Couples may want their wedding to be in weather such as spring and autumn to avoid chance of bad weather in their important day.

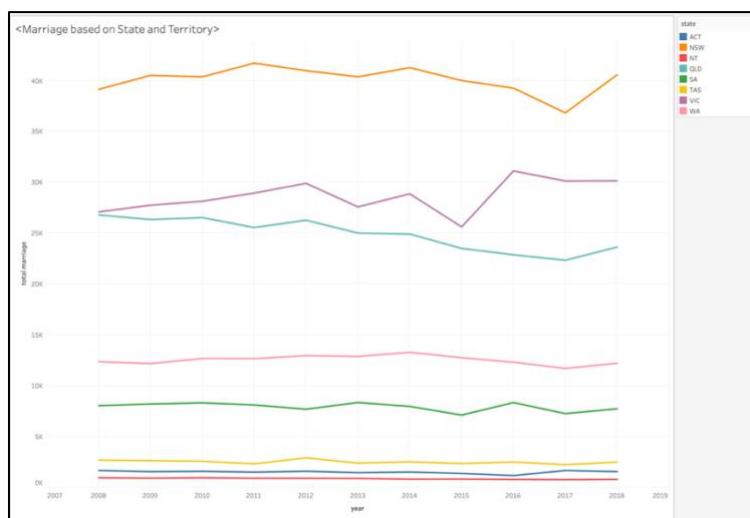


Using Tableau, this graph shows us the average of weddings in a certain date within a month. Different to our previous graph we can see that there are some dates that is highly preferable than others in their month. In their order is 10th of October, 14th of February, and 8th of August. Other than 14th of February, Valentine's Day, the other dates do not have any special meaning to weddings. Though many couples would like to have the same date as their month number, which we can see its reasoning in 10-October (10-10) and 8-August (08-08).

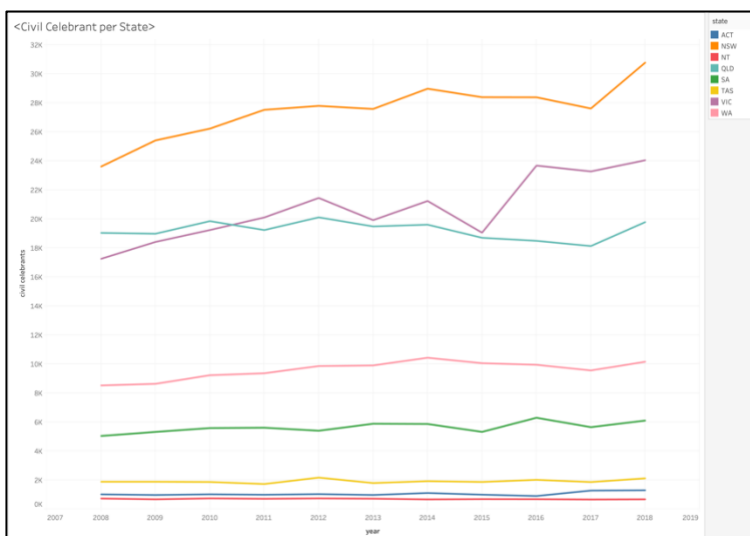
Two noticable dates are less preferred to couples as their date to get married. They are 1st of January and 25th of December. They are indeed special days, yet they are not preferred for date of marriage. This may be caused by their guest availability. Many couples would have to consider when can their guest attend. If they have a wedding, they would want guest to be available and show up. Many would like to spend time with their families in those public holidays. However, having a wedding in those dates is still possible. That is why the number is not zero.

It is also interesting to point out how the curve moves along similarly to how the average marriage occurrence by month graph is shaped. It seems that as the season changes the preference for each date of the month also changes overtime. For an annotated detail of the graph above please see Appendix 8.

How does marriage rate differ by location (States and Territory) and time?

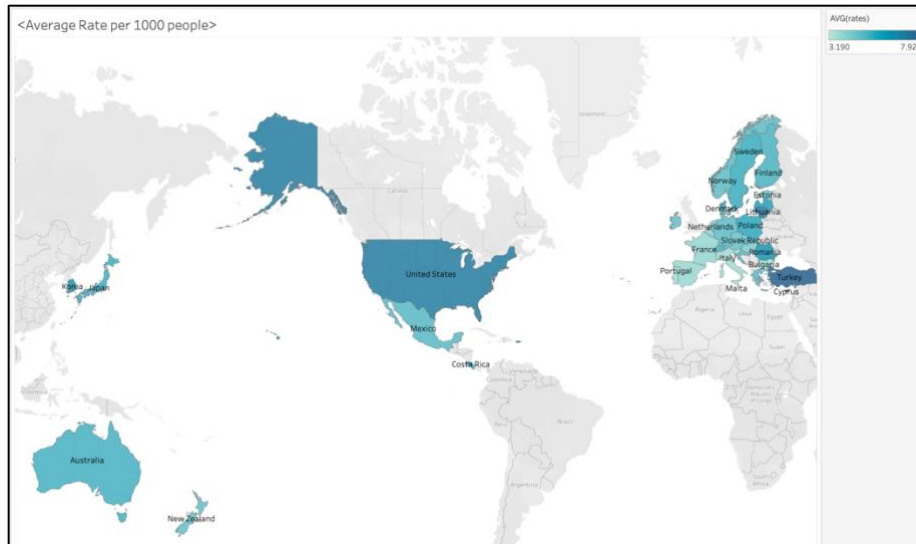


Using Tableau, we present a line representation of the number of marriages within the states and territory. Northern Territory has the lowest of all time and New South Wales have the highest of all time. Overall, there seems to be no changes to the number of marriages within the states. However, Queensland seems to have a decreasing line over the years. It is the most noticeable of tall states. In 2017, there seems to be a slight decrease, although for NSW it is quite significant, yet ACT seems to have an increasing spike. Other than a few, there seems no common spikes that is related to all the states and territories.



As previous analysis, civil celebrant number has increased in the past few years. We see by state that, though sometimes decrease, their movement seems to overall increase. An interesting significant increase in VIC 2015 to 2016 and it is not followed by a significant decrease means that there is a change in behaviour in VIC. Though not as much another significant increase is in NSW from 2017 to 2018. The movement of these lines are very similar to the movement of total marriage above. This may be because they are used in most of the weddings in the states and territories, making their movement very similar.

How does marriage rate differ by location (Australia and Other Countries) and time?



Visual representation of average rates per 1000 people. Australia is deemed to be in medium level as Turkey has the darker colors (OECD, n.d.). Since it is an average, there is not much to see other than how the marriage rates compared to one another. From the line graph in Appendix 4, there seems to be no trend going on that affect all the countries.

We then decide to ask ourselves which countries are have their growth increased and which one decrease from 2008 to 2017. Using RStudio, we calculated the change in rate by subtracting 2017 rate with 2008 rate. Further detail about the data can be seen in Appendix 6. In the result, Australia's rate is negative with -0.9, which is in accordance to our previous analysis. Hungary has the highest growth since 2008 with 1.2 and Turkey has the lowest growth with -1.9 rate change. This is interesting since Turkey has the highest average of 7.92 and Slovenia with the lowest of 3.19. Australia has an average of 5.18.

CONCLUSION

There seems to be no trend in marriages nor divorces that is happening these past few years, contrary to the common folk beliefs. They both fluctuate throughout the years and have no correlation to one another. Marriage factors have changed, though slightly, throughout the years. Less people have preferred marriage by ministry of religion. Also, marriages between people from different birth countries have increased, though marriages between Australians still dominate it is to be expected as out data is from Australia. Males in Australia are more likely to marry in a later age than females. Both are likely to have their first marriage in their late twenties. Throughout the 31 dates available throughout the year, the 13th day is when couples want their wedding the least. Couples may want to avoid bad weather for their wedding thus choosing a month where the weather is similar to spring or autumn is preferable. NSW is the state with the highest number of marriages in Australia. Compared to other countries, Australia has a medium level marriage rate per 1000.

REFLECTION

If I had more time, I would want to compare the number of marriages to the number of education and unemployment. I would also want to compare to questionnaires from people of different ages

thinking what the perfect time and method would to have a marriage to see if there is a difference in the way 20th century compared to 90th century think.


In my original proposal I wanted to compare Australia marriage over the past few years to other countries. However, there was a lack of resources to find necessary crude number for the countries. I can only get their marriage rates per 1000 people and none other to compare further about the factors.

BIBLIOGRAPHY

- Australian Bureau of Statistics. (2020). Marriages and Divorces, Australia - Past Releases. Retrieved from <https://www.abs.gov.au/AUSSTATS/abs@.nsf/second+level+view?ReadForm&prodno=3310.0&viewtitle=Marriages%20and%20Divorces,%20Australia~2018~Latest~27/11/2019&tabname=Past%20Future%20Issues&prodno=3310.0&issue=2018&num=&view=&>. Retrieved 17 September 2020
- Bob. (2018). Calculating Rate of Change. Retrieved from <https://stackoverflow.com/questions/48533631/calculating-rate-of-change>
- Easy Wedding. (n.d.). Religious or civil? Which type of wedding ceremony suits you? Retrieved from <https://www.easyweddings.com.au/articles/religious-or-civil-which-type-of-ceremony-suits-you/>
- OECD. (n.d.). OECD Family Database. Retrieved from <http://www.oecd.org/els/family/database.htm>
- STHDA. (n.d.). Correlation Test Between Two Variables in R. Retrieved from <http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r>

APPENDIX

Appendix 1: Example data viewed in Excel before wrangling

<div>  <h1>Australian Bureau of Statistics</h1> </div>								
3310.0 Marriages and Divorces, Australia, 2018								
Released at 11.30am (Canberra time) 27 November, 2019								
Table 1 Selected marriage indicators, 1998, 2008, 2014-2018(a)(b)								
		1998	2008	2014	2015	2016	2017(i)	2018(j)
ALL MARRIAGES								
Total marriages registered	no. #####	118,755	121,197	113,595	118,401	112,954	119,188	
Crude marriage rate(c)	rate	5.9	5.5	5.2	4.8	4.9	4.6	4.8
Previous marital status(d)								
First marriage both partners	no.	73,743	82,849	87,811	81,715	85,559	81,511	85,954
First marriage one partner	no.	20,705	19,969	19,439	18,542	18,913	17,967	19,584
Remarriage both partners	no.	16,150	15,941	13,950	13,339	13,928	13,475	13,645
Marriage celebrant(e)(f)								
Ministers of religion								
Number	no.	55,807	41,537	31,247	28,409	27,893	24,880	24,178
Proportion	%	50.5	35.0	25.8	25.0	23.6	22.0	20.3
Civil Celebrants								
Number	no.	54,791	77,151	89,891	85,124	90,451	88,061	95,003
Proportion	%	49.5	65.0	74.2	74.9	76.4	78.0	79.7
Relative country of birth of couple(g)								
Both born in Australia	no.	70,876	70,519	66,074	61,565	64,552	60,336	64,353
Both born in the same overseas country	no.	8,890	12,461	16,188	15,762	16,439	16,388	16,724
Born in different countries	no.	30,832	35,628	38,408	36,240	37,374	36,208	38,062
MALES								
Age-specific marriage rates(h)								
Age group (years)								
16-19	rate	1.0	0.8	0.7	0.7	0.6	0.6	0.5
20-24	rate	26.2	17.3	14.3	12.8	12.3	11.3	11.0
25-29	rate	50.8	45.0	41.4	37.9	39.0	35.5	34.9
30-34	rate	31.0	36.6	35.2	32.3	33.6	31.6	32.4
35-39	rate	16.4	20.1	18.4	17.2	18.0	17.1	17.6
40-44	rate	9.8	11.4	11.0	10.5	10.6	10.1	10.3
45-49	rate	7.5	8.1	7.6	7.6	7.9	7.6	8.3
50 and over	rate	3.5	3.7	3.4	3.3	3.5	3.3	3.8
Previous marital status								
Never married	no.	53,972	91,845	96,620	89,826	93,897	89,586	94,120
Widowed	no.	1,162	1,874	1,584	1,531	1,610	1,472	1,552
Divorced	no.	19,240	25,033	22,992	22,246	22,898	21,897	22,422
Total(a)	no.	74,374	118,755	121,197	113,595	118,401	112,954	118,101
Median age at marriage								
Never married	years	27.9	29.6	30.0	30.1	30.3	30.4	30.7
Widowed	years	61.4	62.3	63.7	64.3	64.2	63.7	64.1
Divorced	years	42.0	46.0	47.0	47.4	47.9	47.8	48.3
Total	years	29.8	31.6	31.5	31.7	31.9	32.0	32.4
FEMALES								
Age-specific marriage rates(h)								
Age group (years)								
16-19	rate	5.2	3.5	2.8	2.4	2.2	2.0	1.9
20-24	rate	44.0	30.5	25.0	21.1	21.4	19.3	18.5
25-29	rate	50.2	52.1	48.9	44.0	46.2	41.8	42.4
30-34	rate	24.4	30.9	30.7	28.6	30.0	28.1	29.2
35-39	rate	12.1	15.3	14.2	13.8	14.0	13.6	14.4

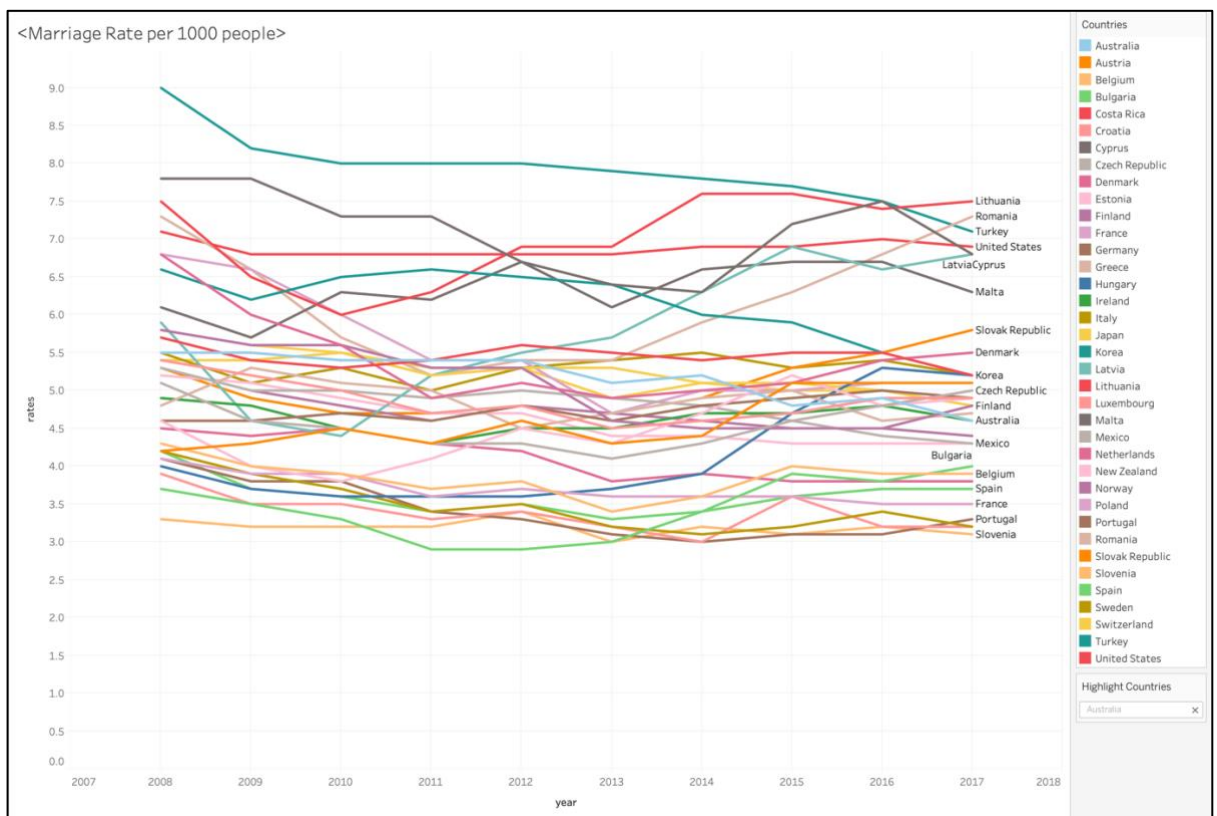
Appendix 2: Example step 1 result of Data Wrangling

	1998	2008	2014	2015	2016	2017	2018
total marriage	110598	118755	121197	113595	118401	112954	119188
first marriage both partners	73743	82849	87811	81715	85559	81511	85954
first marriage one partner	20705	19969	19439	18542	18913	17967	19584
remarriage both partners	16150	15941	13950	13339	13928	13475	13645
ministers of religion	55807	41537	31247	28409	27893	24880	24178
civil celebrants	54791	77151	89891	85124	90451	88061	95003
both born in australia	70876	70519	66074	61565	64552	60336	64353
both born in same overseas country	8890	12461	16188	15762	16439	16388	16724
born in different countries	30832	35628	38408	36240	37374	36208	38062

Appendix 3: 2018 File I - viewed using head() function in RStudio

Description <chr>	1998 <dbl>	2008 <dbl>	2014 <dbl>	2015 <dbl>	2016 <dbl>	2017 <dbl>	2018 <dbl>
total marriage	110598	118755	121197	113595	118401	112954	119188
first marriage both partners	73743	82849	87811	81715	85559	81511	85954
first marriage one partner	20705	19969	19439	18542	18913	17967	19584
remarriage both partners	16150	15941	13950	13339	13928	13475	13645
ministers of religion	55807	41537	31247	28409	27893	24880	24178
civil celebrants	54791	77151	89891	85124	90451	88061	95003

Appendix 4: Line graph for each countries marriage rate per 1000 people



Appendix 6: Change in rate of change from 2008 to 2017

countries	2017	2008	change
Australia	5.5	4.6	-0.9
Austria	4.2	5.1	0.9
Belgium	4.3	3.9	-0.4
Czech Republic	5.1	5.0	-0.1
Denmark	6.8	5.5	-1.3
Estonia	4.6	4.9	0.3
Finland	5.8	4.8	-1.0
France	4.1	3.5	-0.6
Germany	4.6	4.9	0.3
Greece	4.8	4.7	-0.1
Hungary	4.0	5.2	1.2
Ireland	4.9	4.6	-0.3
Italy	4.2	3.2	-1.0
Japan	5.8	4.9	-0.9
Korea	6.6	5.2	-1.4
Latvia	5.9	6.8	0.9
Lithuania	7.5	7.5	0.0
Luxembourg	3.9	3.2	-0.7
Mexico	5.3	4.3	-1.0
Netherlands	4.5	3.8	-0.7
New Zealand	5.2	4.3	-0.9
Norway	5.3	4.4	-0.9
Poland	6.8	5.1	-1.7
Portugal	4.1	3.3	-0.8
Slovak Republic	5.3	5.8	0.5
Slovenia	3.3	3.1	-0.2
Spain	4.2	3.7	-0.5
Sweden	5.5	5.2	-0.3
Switzerland	5.4	4.8	-0.6
Turkey	9.0	7.1	-1.9
United States	7.1	6.9	-0.2
Costa Rica	5.7	5.2	-0.5
Bulgaria	3.7	4.0	0.3
Croatia	5.4	4.9	-0.5
Cyprus	7.8	6.8	-1.0
Malta	6.1	6.3	0.2
Romania	7.3	7.3	0.0

Appendix 7: Average marriage rate per 1000

Countries	average
Australia	5.18
Austria	4.59
Belgium	3.85
Bulgaria	3.44
Costa Rica	5.45
Croatia	4.87
Cyprus	7.11
Czech Republic	4.56
Denmark	5.43
Estonia	4.49
Finland	5.05
France	3.70
Germany	4.75
Greece	4.86
Hungary	4.13
Ireland	4.63
Italy	3.48
Japan	5.28
Korea	6.14
Latvia	5.79
Lithuania	7.02
Luxembourg	3.38
Malta	6.34
Mexico	4.82
Netherlands	4.10
New Zealand	4.63
Norway	4.72
Poland	5.51
Portugal	3.40
Romania	6.19
Slovak Republic	5.06
Slovenia	3.19
Spain	3.61
Sweden	5.30
Switzerland	5.17
Turkey	7.92
United States	6.88

Appendix 8: Average Marriage Occurrence by Day of the Year – Annotated

