Problem Set

Tim Hogan

10/1/2019

Problem 1

How many continents are included?

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

How many countries are included?

```
## [1] 142
```

How many countries per continent?

```
## # A tibble: 5 x 2
##   continent `Number of Countries`
##   <fct>                     <int>
## 1 Africa                       52
## 2 Americas                     25
## 3 Asia                         33
## 4 Europe                       30
## 5 Oceania                       2
```

Creating the report. I chose to narrow the report to 2007, and calculating the total population and GDP per capity per continent.

Continent	Total Population	GDP per Capita
Africa	929539692	160629.70
Americas	898871184	275075.79
Asia	3811953827	411609.89
Europe	586098529	751634.45
Oceania	24549947	59620.38

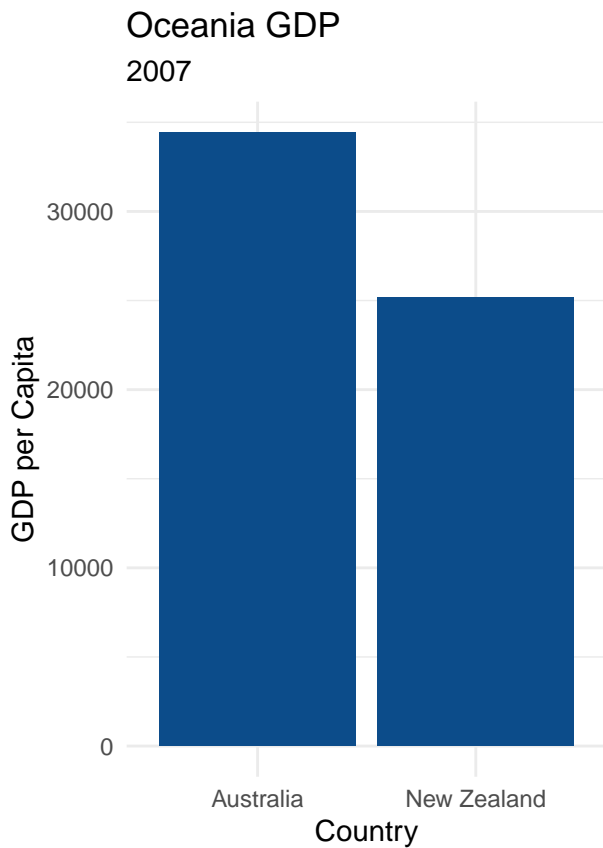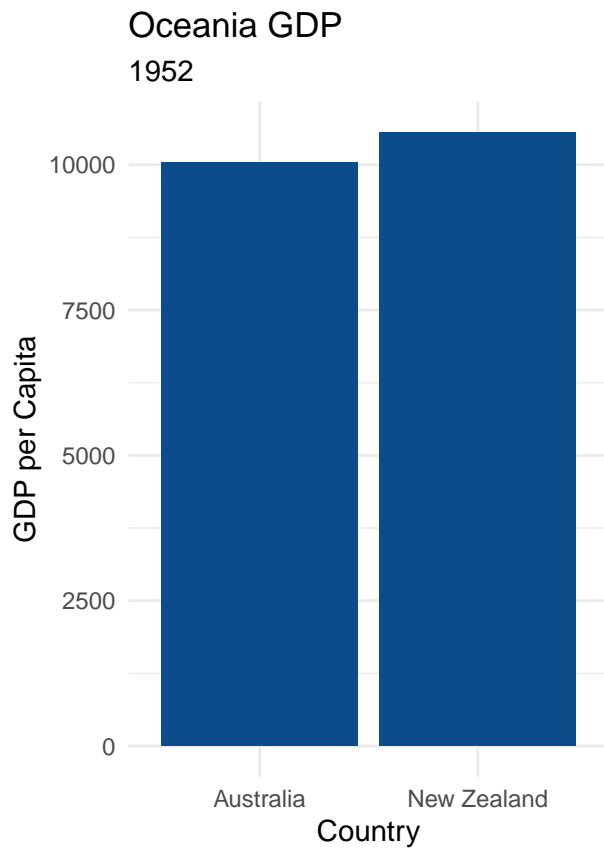**Table summarizing GDP per capita for countries in each year.**

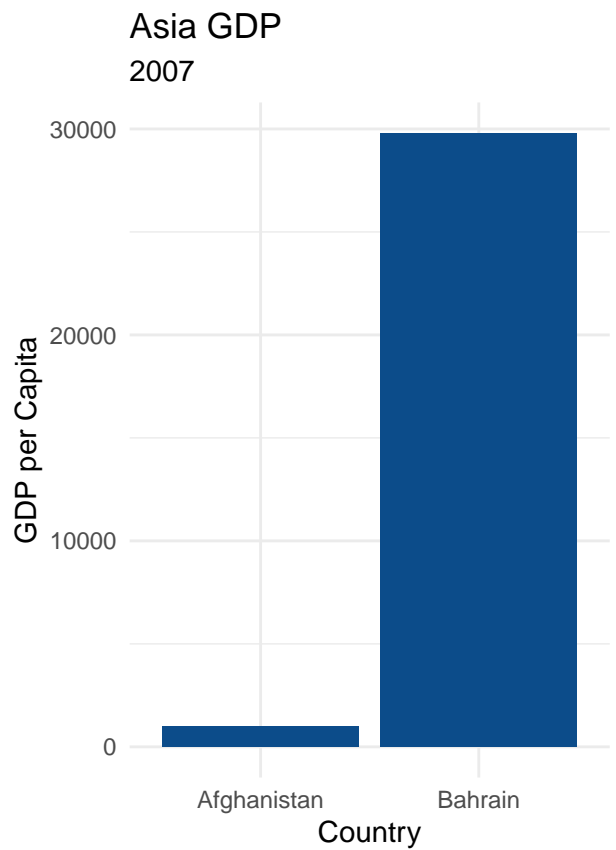This table will only display the data for two countries from each continent.
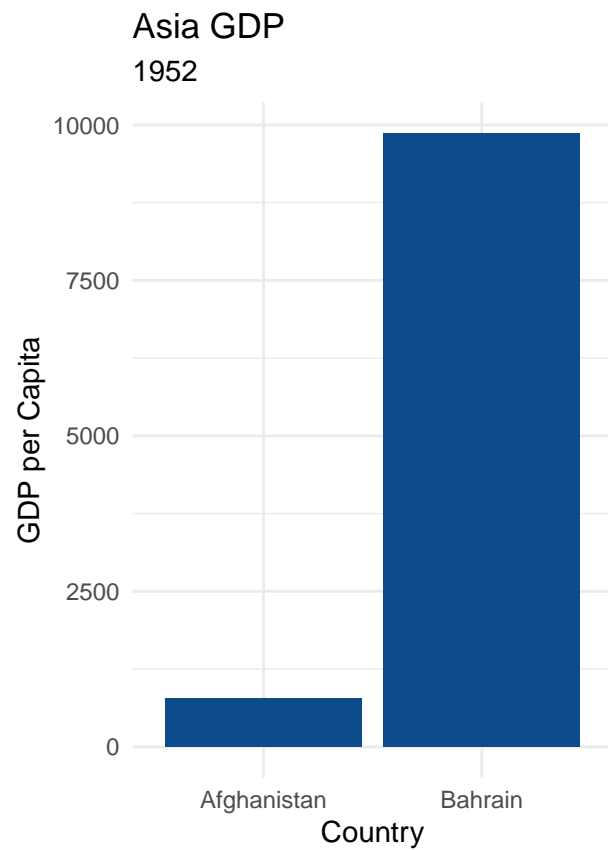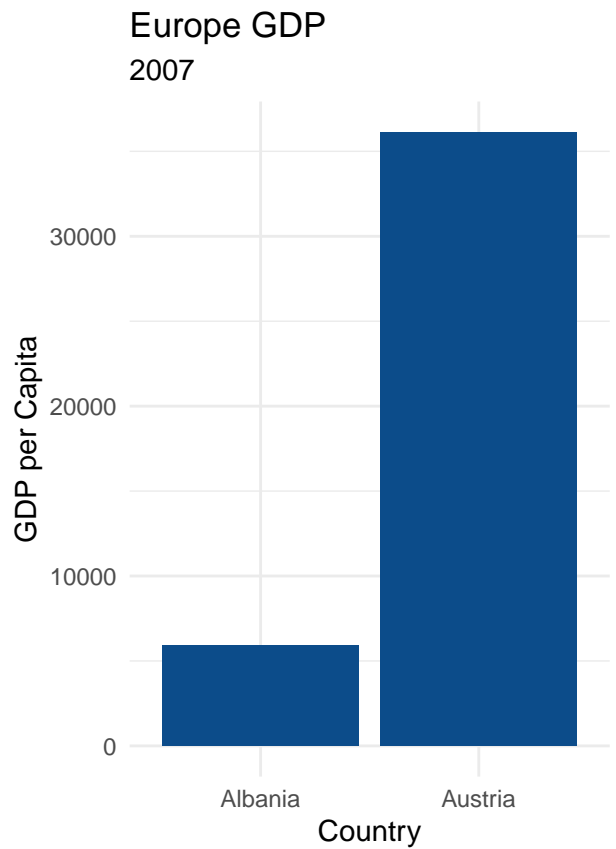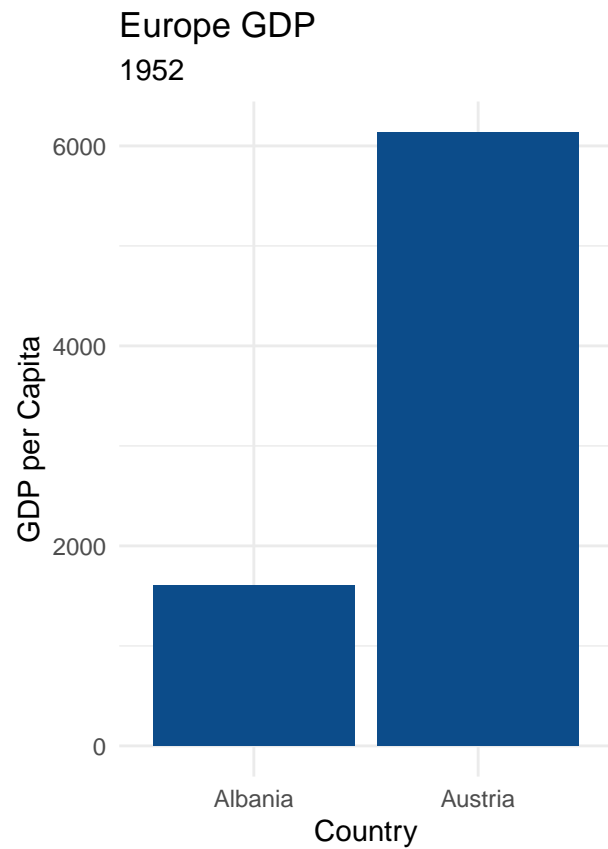
The reason for this is that there are only two countries in Oceania. For the neatness and simple display of the table, I will only show two of each.

Country	Continent	1952 GDP	2007 GDP
Algeria	Africa	2449.0	6223.4
Angola	Africa	3520.6	4797.2
Argentina	Americas	5911.3	12779.4
Bolivia	Americas	2677.3	3822.1
Afghanistan	Asia	779.4	974.6
Bahrain	Asia	9867.1	29796.0
Albania	Europe	1601.1	5937.0
Austria	Europe	6137.1	36126.5
Australia	Oceania	10039.6	34435.4
New Zealand	Oceania	10556.6	25185.0

Plots for the above.

To reflect the table, I will only graph two countries per plot.

Oceania GDP
1952

Oceania GDP
2007

Europe GDP
1952

Europe GDP
2007

Asia GDP
1952

Asia GDP
2007

America GDP
1952

America GDP
2007

Africa GDP
1952

Africa GDP
2007
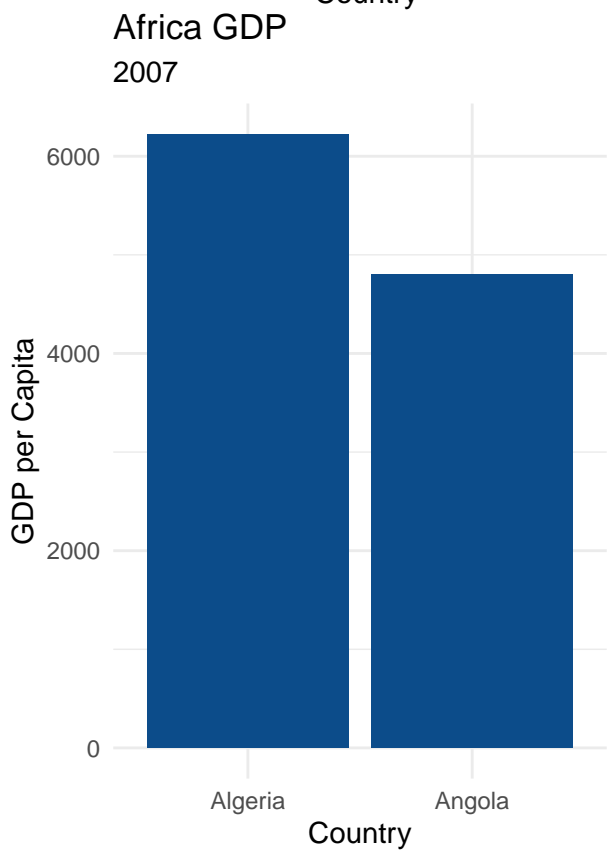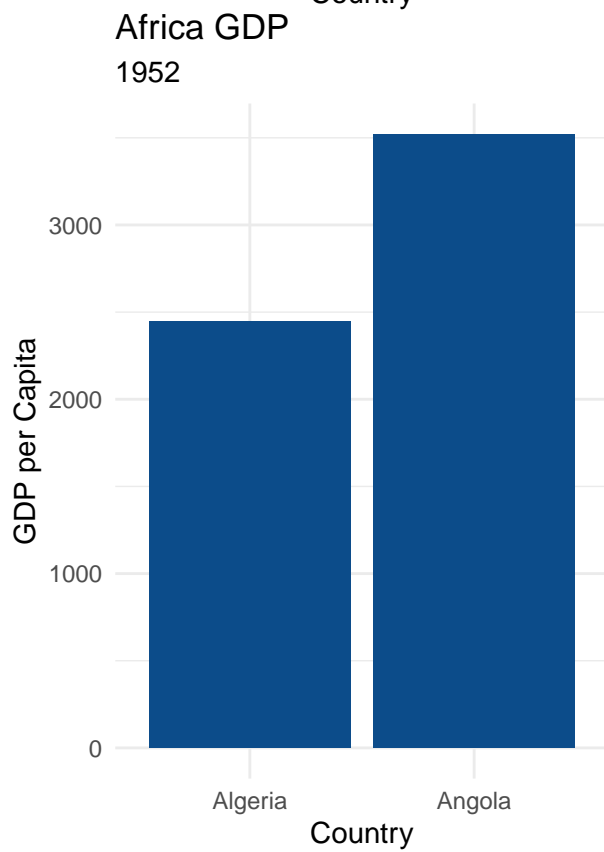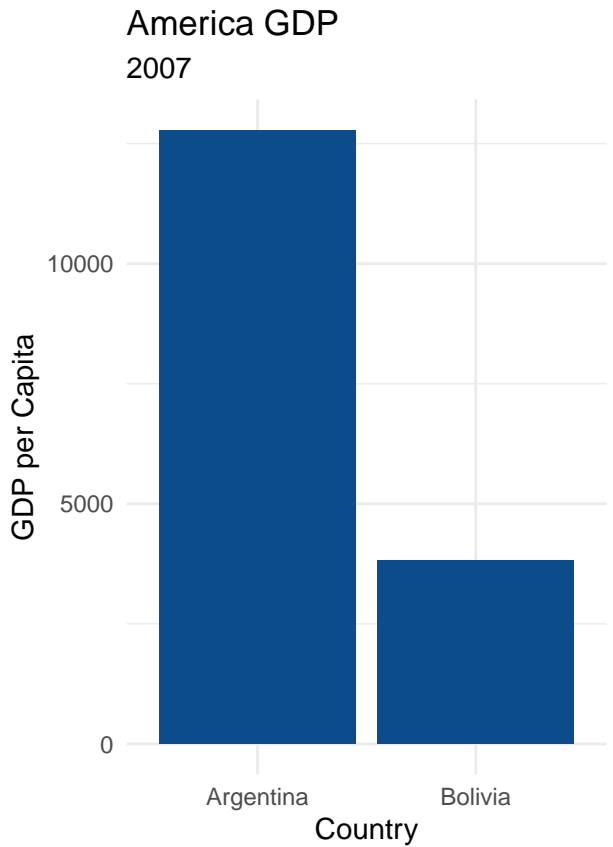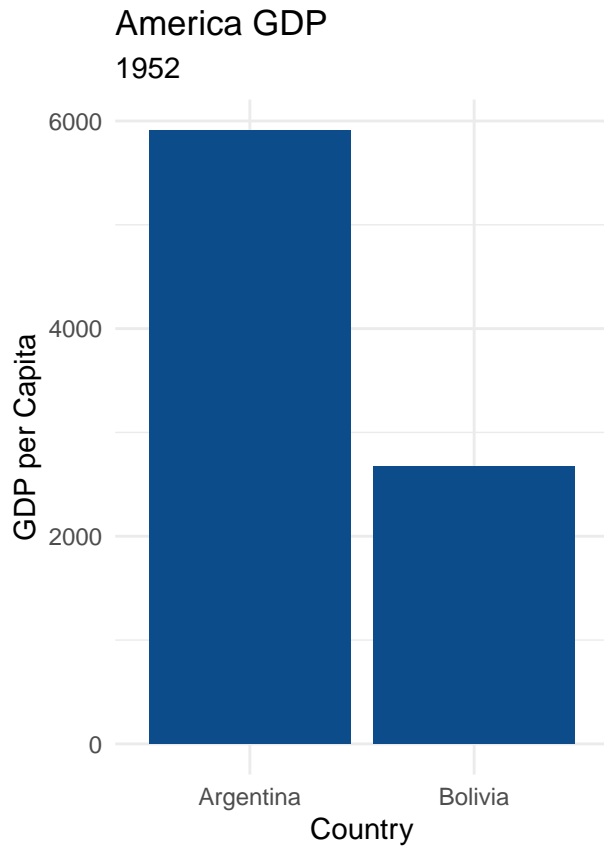
Negative Population Growth

```r
gaptest <- gapminder %>% group_by(country) %>%
  mutate(growth = pop - lag(pop, default=0))

gapdecrease <- gaptest %>% filter(growth <0)

kable(gapdecrease[,c("country","year","growth")], booktabs=TRUE, digits = 2, col.names = c("Country", "Y
  kable_styling(font_size=18)
```

| Country	Year	Difference
Afghanistan	1982	-1998556
Bosnia and Herzegovina	1992	-82964
Bosnia and Herzegovina	1997	-649013
Bulgaria	1992	-313452
Bulgaria	1997	-592449
Bulgaria	2002	-404258
Bulgaria	2007	-338941
Cambodia	1977	-471999
Croatia	1997	-49418
Czech Republic	1997	-14995
Czech Republic	2002	-44412
Czech Republic	2007	-27551
Equatorial Guinea	1977	-84928
Germany	1977	-556315
Germany	1987	-616968
Guinea-Bissau	1967	-26533
Hungary	1987	-92795
Hungary	1992	-264056
Hungary	1997	-104000
Hungary	2002	-161371
Hungary	2007	-127205
Ireland	1957	-73936
Ireland	1962	-48220
Kuwait	1992	-473392
Lebanon	1982	-28911
Lesotho	2007	-34123
Liberia	1992	-356440
Montenegro	2007	-35494
Poland	2002	-28981
Poland	2007	-107735
Portugal	1972	-132550
Romania	1997	-234569
Romania	2002	-158121

Highest Growth in GDP

I will interpret this as the highest growth rate in GDP in a 5 year period.

Country	End of 5 year period	Change in GDP
Kuwait	1972	28452.98
Kuwait	2007	12196.88
Libya	1967	12015.72
Singapore	2007	11120.07
Gabon	1977	10343.62
Ireland	2002	9555.10
Hong Kong, China	2007	9515.96
Saudi Arabia	1977	9330.33
Singapore	1997	8749.59
Saudi Arabia	1972	7934.38

Problem 2

Gender combinations and age groups

```
## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival
```
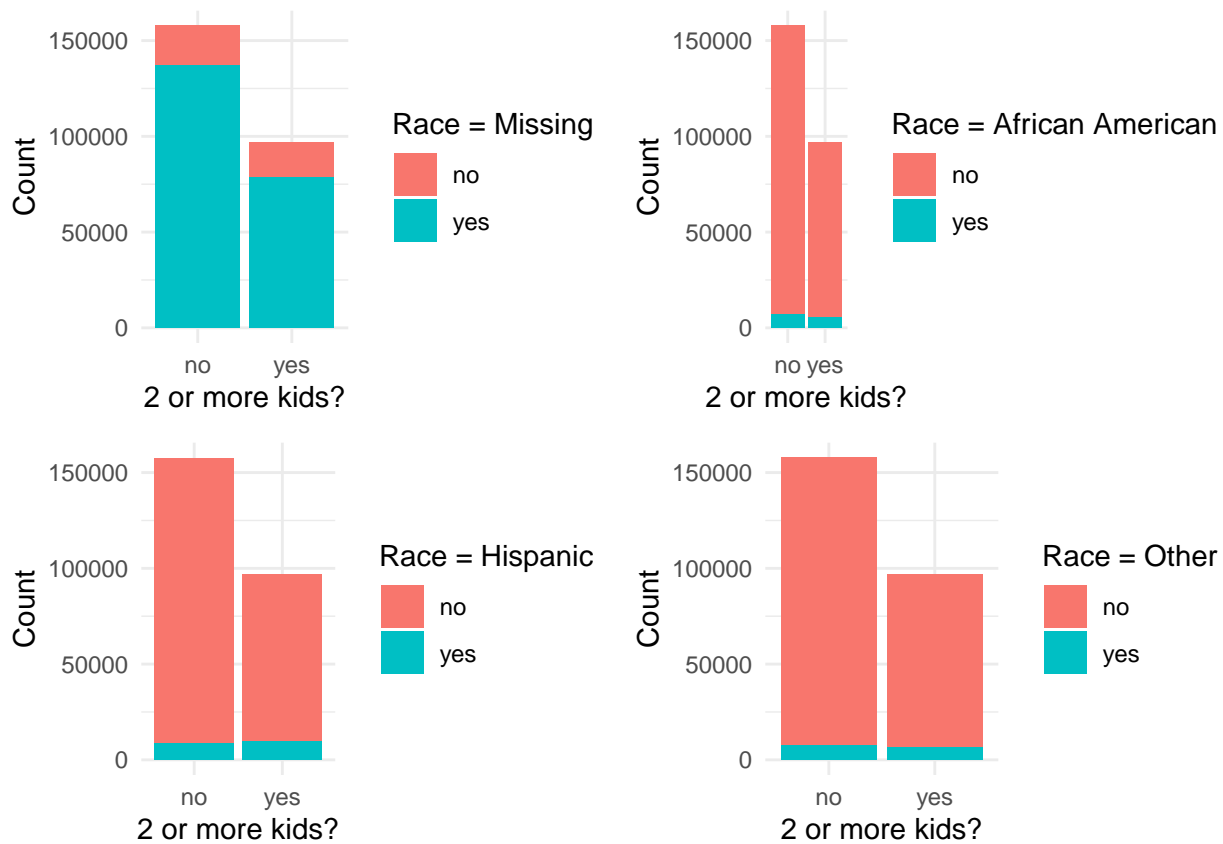
Frequency of Gender Combinatinons of Children
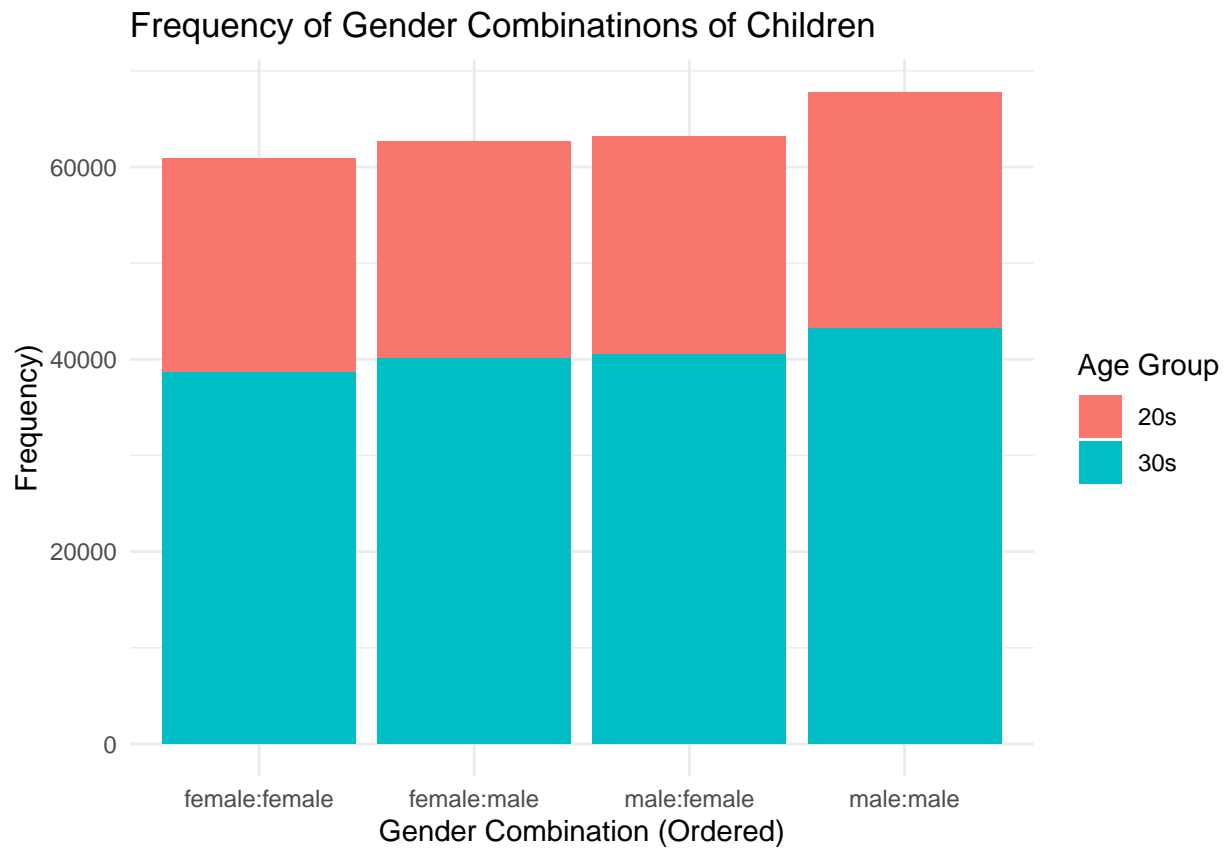
The above plot shows the frequencies of the combinations. We can also see that there are more individuals having more than one children in their 30s than their 20s, leading to higher frequencies overall. However, there does not seem to be a significant difference in their distributions.

Correlation with race

For gender differences, it appears that the "missing" group (which most likely corresponds to white people) frequently has 2 or more kids. Minorities seem to have lower amounts.

Problem 3

Letters and names

Number of e's	Mercs in mtcars	mercs in mpg
25	7	4

These are the values for the number of times "e" appears, the number times "Merc" appears in mtcars, and the number fo times "merc' appears in mpg.

Comparison of MPGs

```
## Warning in mtcars$names == c("Merc 240D", "Merc 230", "Merc 280", "Merc
## 280C", : longer object length is not a multiple of shorter object length
```

Model	City MPG	Highway MPG	Average MPG
mountaineer 4wd	14	17	15.5
mountaineer 4wd	13	19	16.0
mountaineer 4wd	13	19	16.0
mountaineer 4wd	13	17	15.0

	Model	MPG
Merc 240D	Merc 240D	24.4
Merc 230	Merc 230	22.8
Merc 280	Merc 280	19.2
Merc 280C	Merc 280C	17.8
Merc 450SE	Merc 450SE	16.4
Merc 450SL	Merc 450SL	17.3
Merc 450SLC	Merc 450SLC	15.2

Comparing the three numbers, we can make a few assessments: in the mtcars table, Merc 450SLC has the lowest mileage.

Looking at mpg, if we compare city milages to mtcars, all values fall below the mpgs in mtcars.

Looking at highways, they fall in the upper rangewith lower values comparable to the Merc450SL and higher values comparable to the Merc 480,

Looking at the averages, they fall once again on the lower end, the values being comparable to Merc 450SLC and MErc 450SE.

Problem 4

Sampling and displaying the most popular names in each year.

Year	Name	Year	Name	Year	Name	Year	Name
1880	John	1920	Mary	1960	David	2000	Jacob
1880	William	1920	John	1960	Michael	2000	Michael
1880	Mary	1920	William	1960	James	2000	Matthew
1880	James	1920	Robert	1960	John	2000	Joshua
1880	Charles	1920	James	1960	Robert	2000	Emily

Overlap between genders

Names used in the 19th but not the 21st century

Frequencies of Donald, Hilary, Hillary, Joe, and Barrack

Barrack does not appear in the table because it does not appear in the sample.

Frequency of Names
From 1880 to 2017