615 hw2

Shicheng Wang

10/2/2019

Problem 1

How many continents are included in the data set?

```
## # A tibble: 5 x 1
##   continent
##   <fct>
## 1 Asia
## 2 Europe
## 3 Africa
## 4 Americas
## 5 Oceania
```

There are five continents shows in the data.
How many countrys are included? How many countries per continent?

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

```
##
##   Africa Americas     Asia   Europe  Oceania
##      624      300      396      360       24
```

There are 142 countries.
Produce a report

```
## # A tibble: 60 x 4
## # Groups:   continent [5]
##    continent  year   pop_con  gdp_cap
##    <fct>     <int>     <dbl>    <dbl>
##  1 Africa     1952 237640501 0.000274
##  2 Africa     1957 264837738 0.000272
##  3 Africa     1962 296516865 0.000280
##  4 Africa     1967 335289489 0.000318
##  5 Africa     1972 379879541 0.000320
##  6 Africa     1977 433061021 0.000311
##  7 Africa     1982 499348587 0.000258
##  8 Africa     1987 574834110 0.000206
##  9 Africa     1992 659081517 0.000180
## 10 Africa     1997 743832984 0.000166
## # ... with 50 more rows
```

Table 2: Contrast table

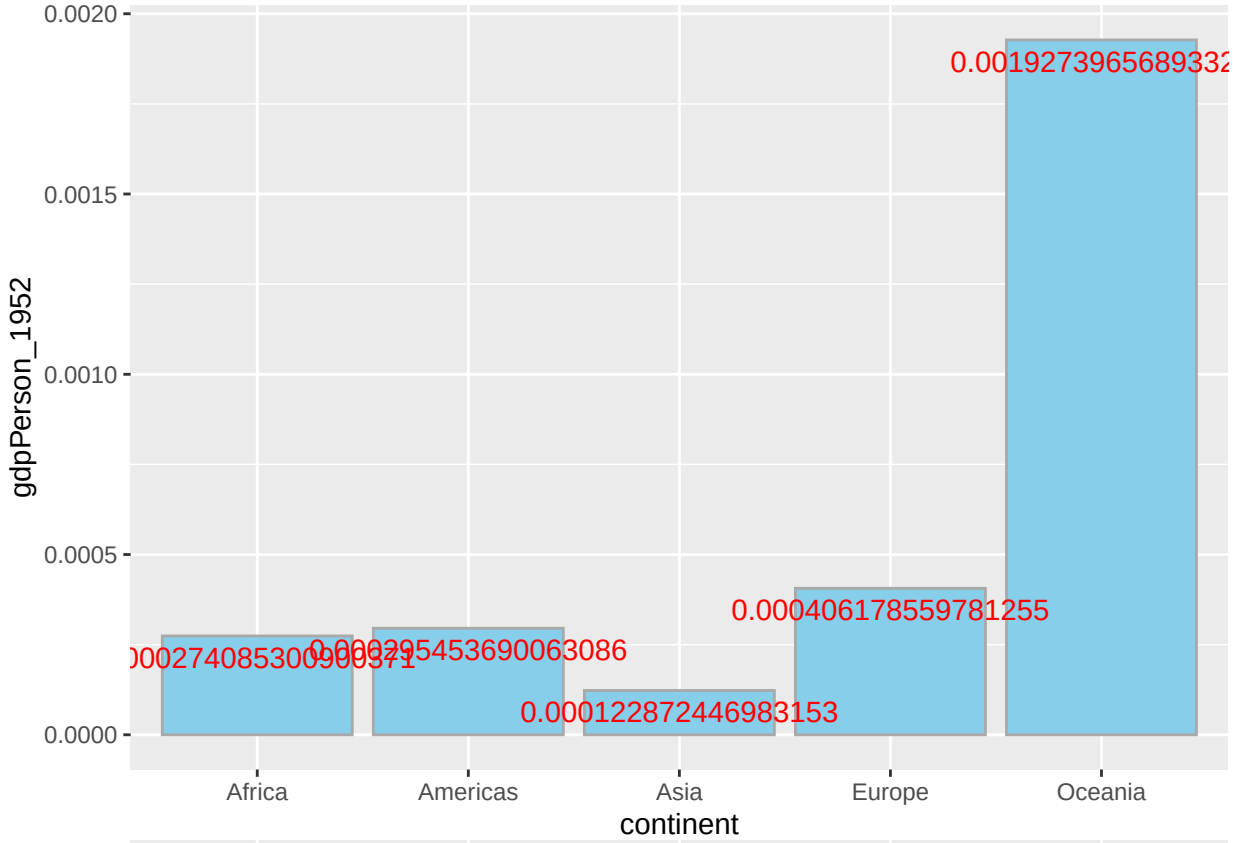
continent	gdpPerson_1952	gdpPerson_2007
Africa	0.0002741	0.0001728
Americas	0.0002955	0.0003060
Asia	0.0001229	0.0001080
Europe	0.0004062	0.0012824
Oceania	0.0019274	0.0024285

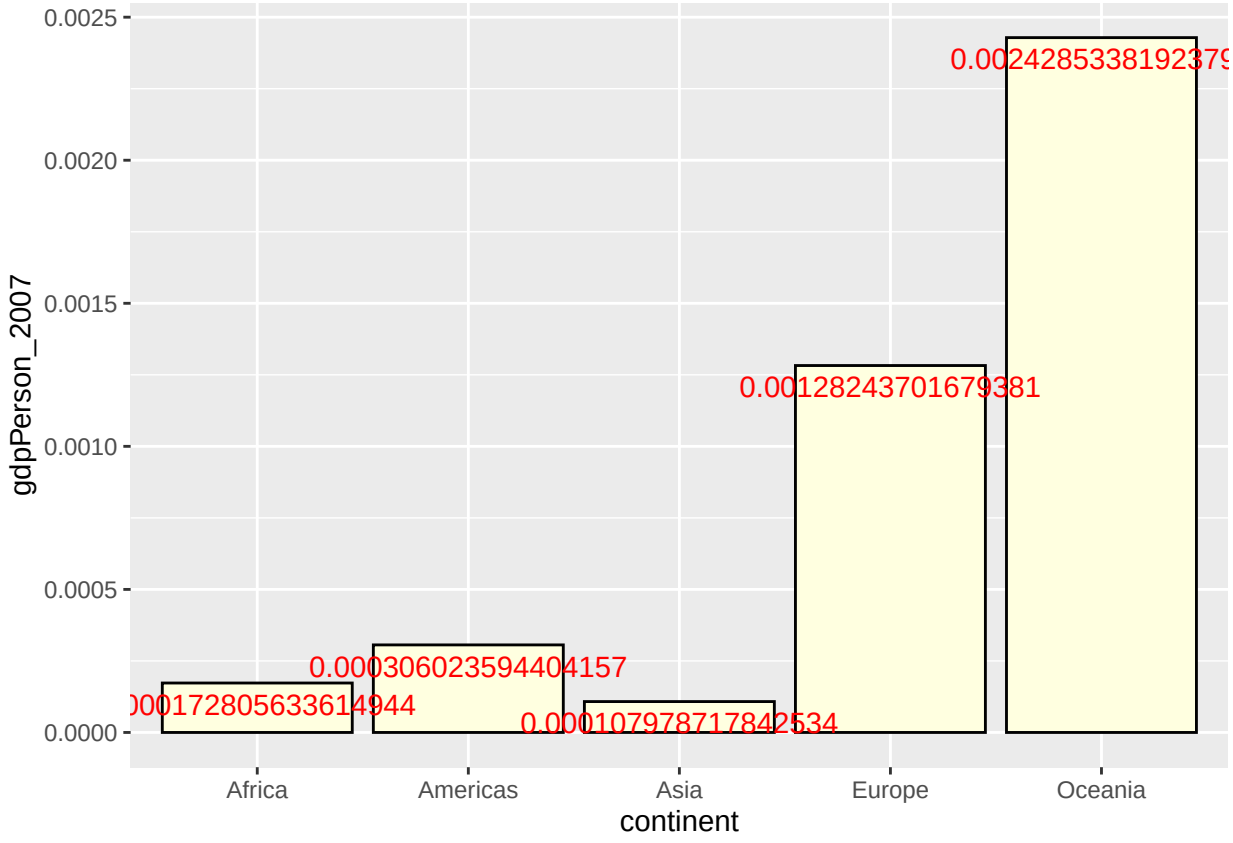
Table 1: Total population and total GDP for each continents

Year	Africa		American		Asia		Europe		Oceania	
continent	year	pop_con	gdp_cap	continent1	year1	pop_con1	gdp_cap1	continent2	year2	pop_con2
Africa	1952	237640501	0.0002741	Americas	1952	345152446	0.0002955	Asia	1952	139535735
Africa	1957	264837738	0.0002720	Americas	1957	386953916	0.0002982	Asia	1957	156278059
Africa	1962	296516865	0.0002803	Americas	1962	433270254	0.0002828	Asia	1962	169635718
Africa	1967	335289489	0.0003180	Americas	1967	480746623	0.0002948	Asia	1967	190566290
Africa	1972	379879541	0.0003203	Americas	1972	529384210	0.0003066	Asia	1972	215097224
Africa	1977	433061021	0.0003105	Americas	1977	578067699	0.0003180	Asia	1977	238451355
Africa	1982	499348587	0.0002584	Americas	1982	630290920	0.0002977	Asia	1982	261013558
Africa	1987	574834110	0.0002065	Americas	1987	682753971	0.0002854	Asia	1987	287122076
Africa	1992	659081517	0.0001800	Americas	1992	739274104	0.0002721	Asia	1992	313329219
Africa	1997	743832984	0.0001663	Americas	1997	796900410	0.0002789	Asia	1997	338328556
Africa	2002	833723916	0.0001621	Americas	2002	849772762	0.0002732	Asia	2002	360180220
Africa	2007	929539692	0.0001728	Americas	2007	898871184	0.0003060	Asia	2007	381195381

Produce a table

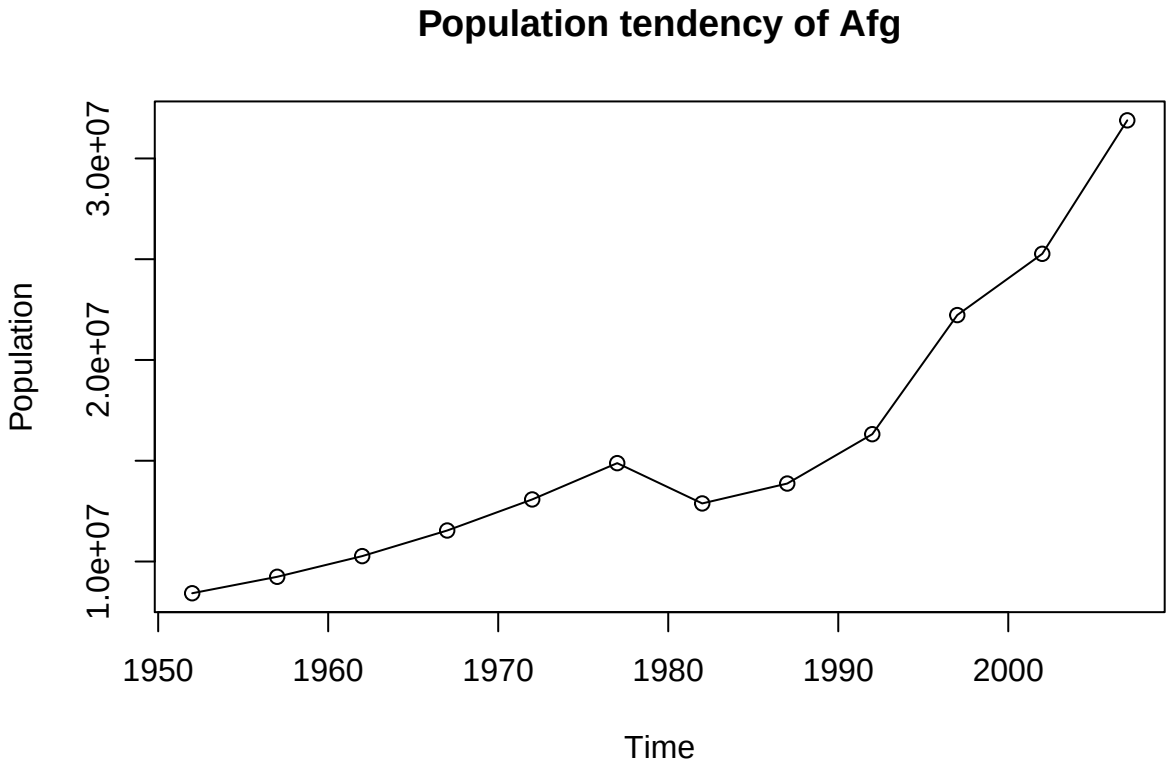
Summary

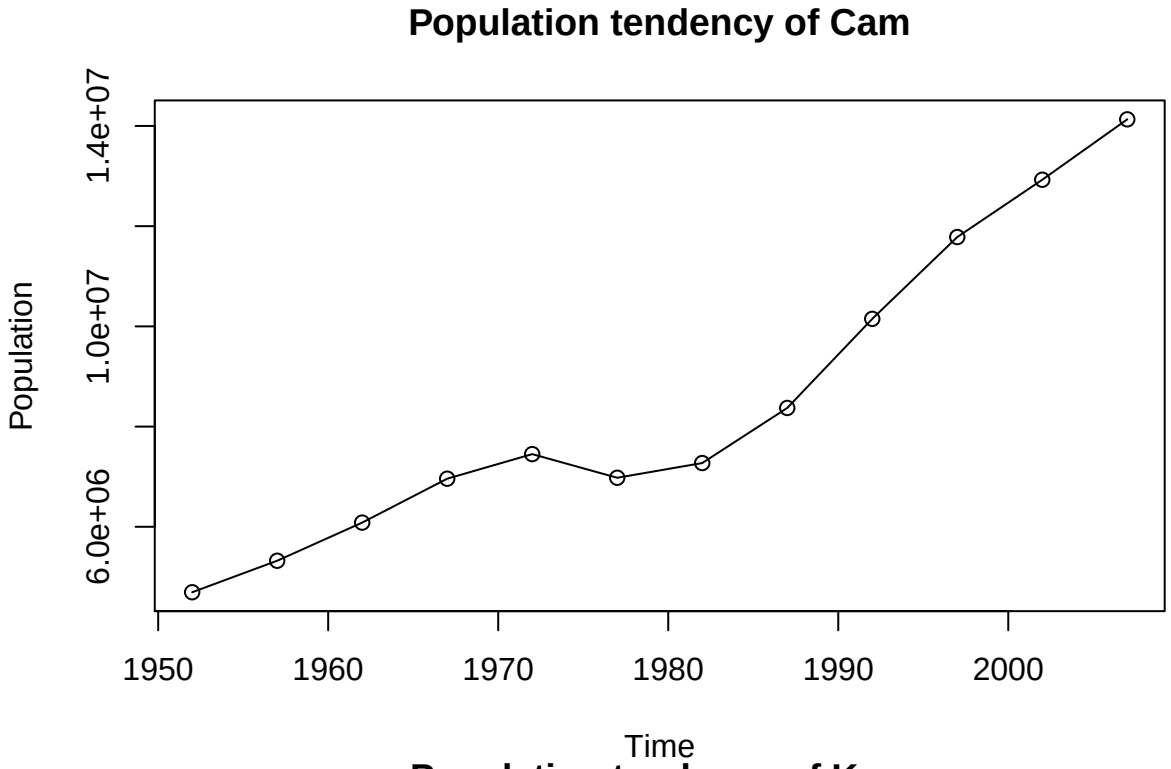


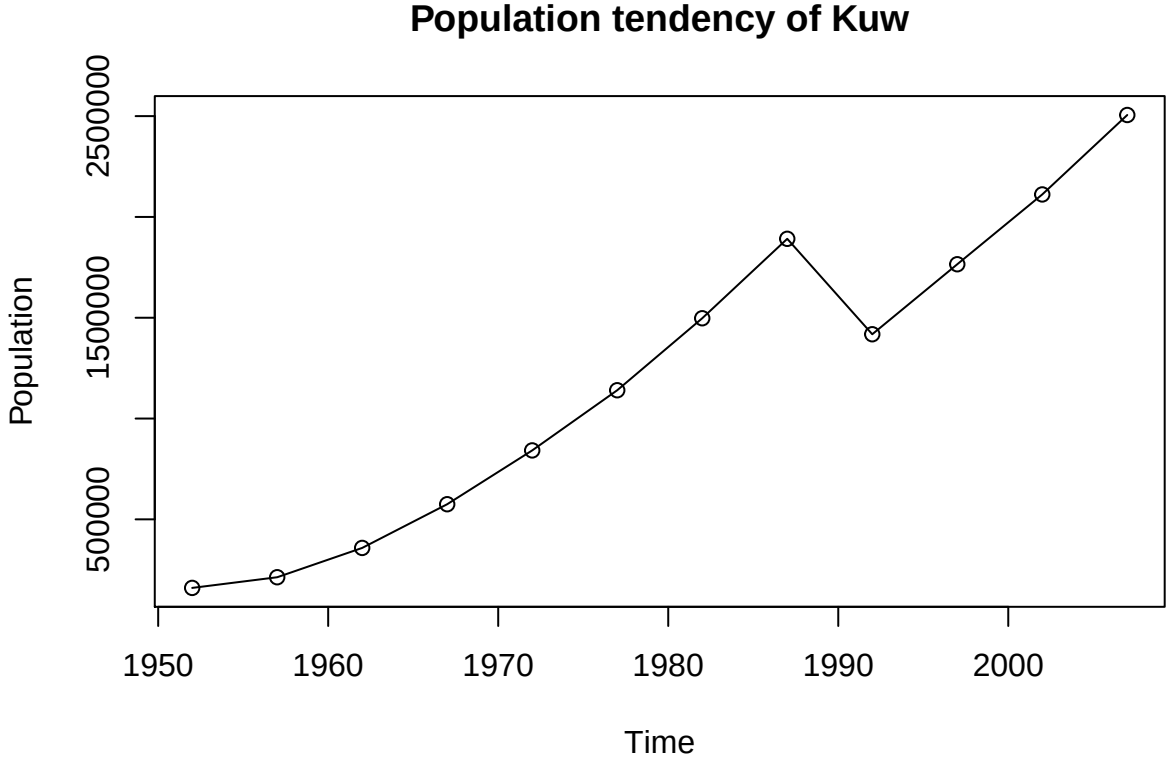


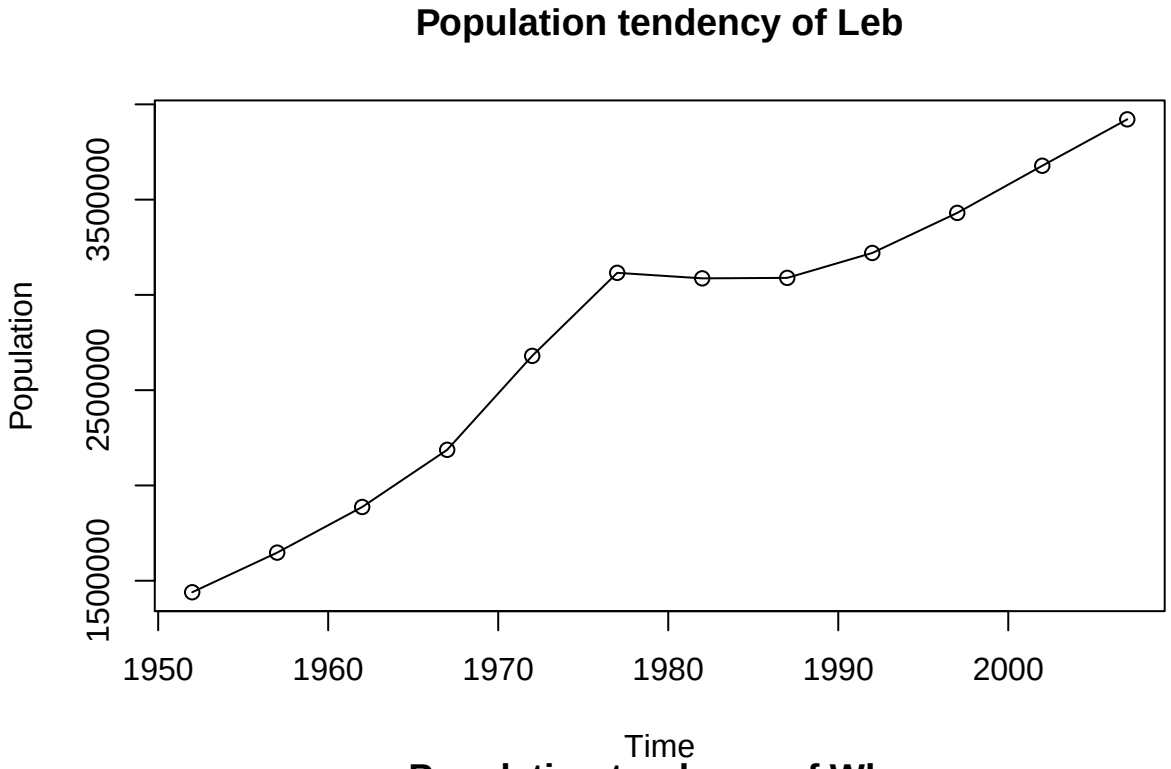
Which countries in the dataset have had periods of negative population growth? Illustrate your answer with a table or plot.

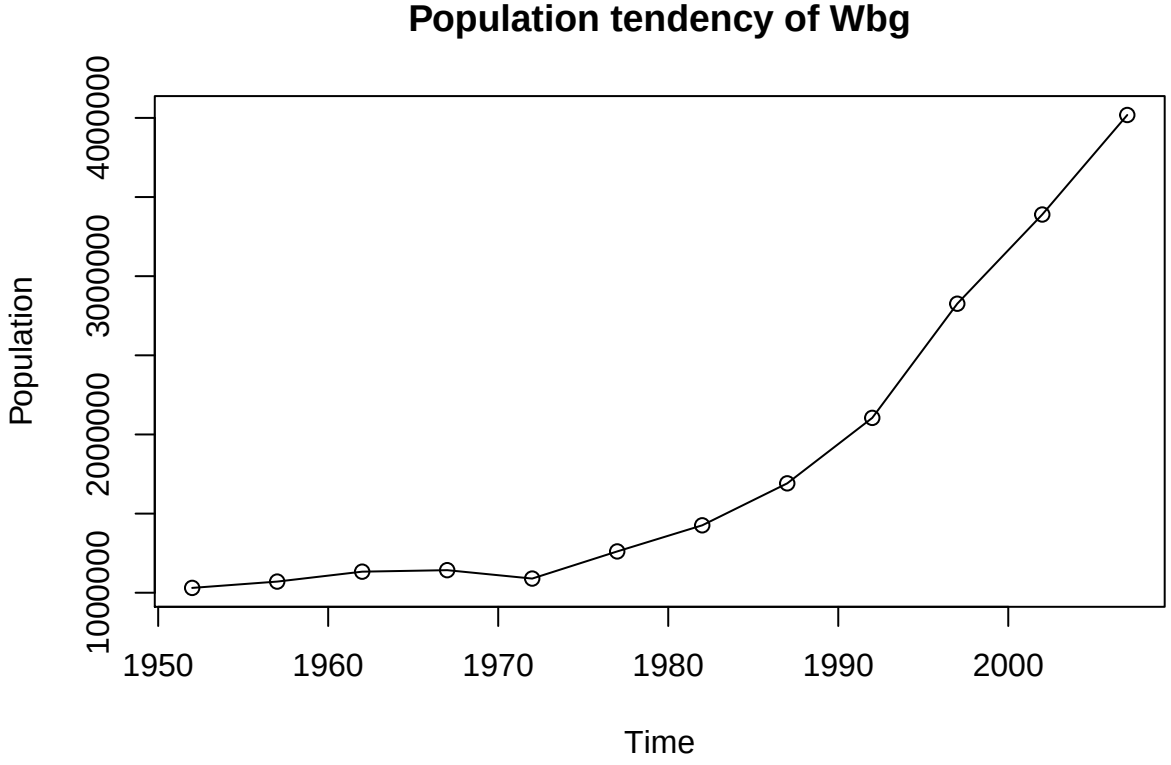
```
## # A tibble: 33 x 1
##    country
##    <fct>
##  1 Afghanistan
##  2 Bahrain
##  3 Bangladesh
##  4 Cambodia
##  5 China
##  6 Hong Kong, China
##  7 India
##  8 Indonesia
##  9 Iran
## 10 Iraq
## # ... with 23 more rows
```




Population tendency of Cam
Population
6.0e+06
1.0e+07
1.4e+07
1950
1960
1970
1980
1990
2000
Time


Population tendency of Kuw
Population
500000
1500000
2500000
1950
1960
1970
1980
1990
2000
Time





The decrease time period of the countries above is:
Afghanistan:1972-1977
Cambodia:1972-1977
Kuwait:1987-1992
Lebanon:1977-1987
West Bank and Gaza:1967-1972

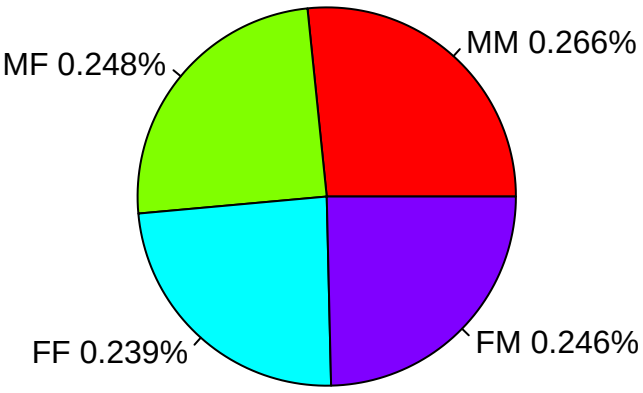
Which countries in the dataset have had the highest rate of growth in per capita GDP?

Illustrate your answer with a table or plot.

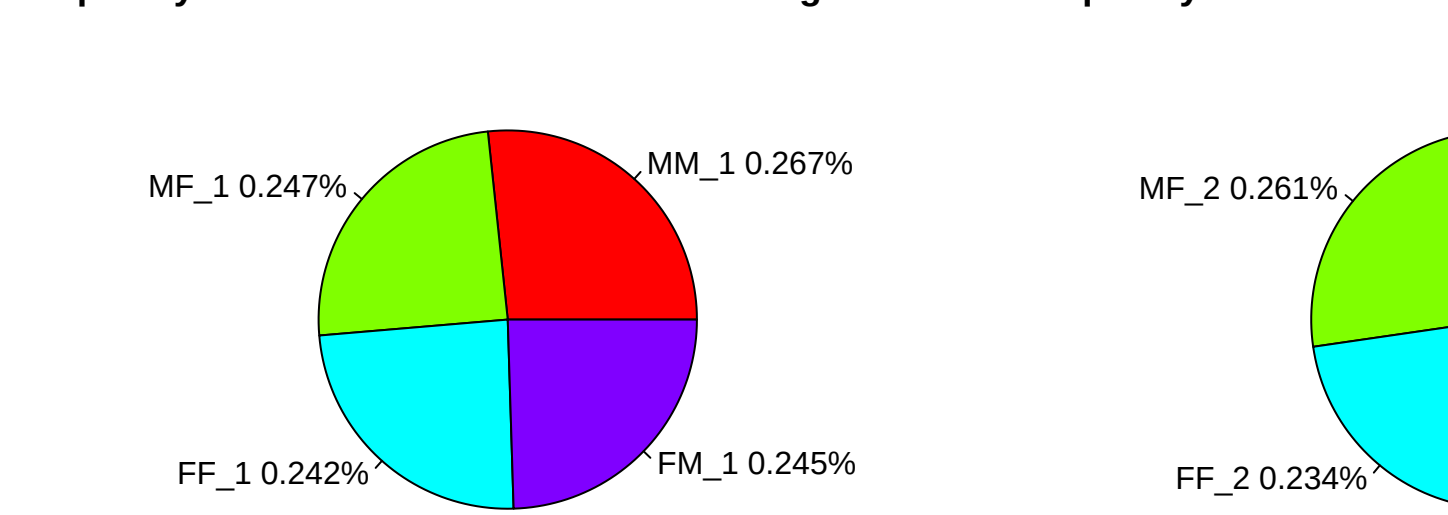
```
## # A tibble: 1 x 7
##   country           continent  year lifeExp    pop gdpPercap  rate
##   <fct>             <fct>     <int>   <dbl>  <int>     <dbl> <dbl>
## 1 Equatorial Guinea Africa     1952    34.5 216964      376.  31.4
```

Problem 2

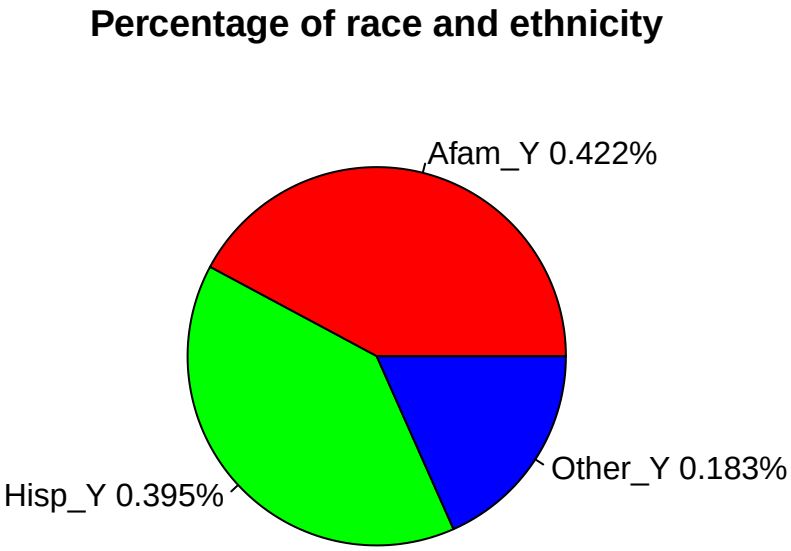
The data for Problem 2 is the Fertility data in the AER package. This data is from the 1980 US Census and is comprised of date on married women aged 21-35 with two or more children. The data report the gender of each woman's first and second child, the woman's race, age, number of weeks worked in 1979, and whether the woman had more than two children.



There are four possible gender combinations for the first two Children. Product a plot the contracts the frequency of these four combinations. Are the frequencies different for women in their 20s and wemen who are older than 29?



Produce a plot that contrasts the frequency of having more than two children by race and ethnicity.



Problem 3

Use the mtcars and mpg datasets.

How many times does the letter "e" occur in mtcars rownames?

```
data(mtcars)
e_number<-sum(str_count(rownames(mtcars),"e"))
e_number
```

```
## [1] 25
```

Table 3: mtcars' Merc

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

Table 4: mpg's Merc

manufacturer	cty	hwy
mercury	14	17
mercury	13	19
mercury	13	19
mercury	13	17

How many cars in mtcars have the brand Merc?

```
## [1] 7
```

How many cars in mpg have the brand("manufacturer" in mpg) Merc?

```
## [1] 4
```

Contrast the mileage data for Merc cars as reported in mtcars and mpg. Use tables, plots, and a short explaination.

Problem 4

Install the babynames package.

Draw a sample of 500,000 rows from the babynames data

```
## # A tibble: 500,000 x 5
##     year sex   name         n       prop
##    <dbl> <chr> <chr>    <int>      <dbl>
##  1  2007 M     Jamel      204 0.0000922
##  2  2002 M     Madison    104 0.0000504
##  3  1915 M     Jonathan    83 0.0000942
##  4  1938 M     Ali          7 0.00000616
##  5  1920 M     Walker     151 0.000137
##  6  1909 F     Ineze        6 0.0000163
##  7  1990 F     Kady       115 0.0000560
##  8  2013 F     Allisen      9 0.00000468
##  9  1998 M     Deveon      17 0.00000839
## 10  1933 F     Zelpha      11 0.0000105
## # ... with 499,990 more rows
```

What names overlap boys and girls?

```
## # A tibble: 51,374 x 2
##    name       lap
##    <chr>    <int>
##  1 Aaban        2
##  2 Aadam        6
##  3 Aadan        5
##  4 Aadarsh      5
##  5 Aaden        5
##  6 Aadhav       4
##  7 Aadhavan     3
##  8 Aadhi        2
##  9 Aadhira      2
## 10 Aadhya       3
## # ... with 51,364 more rows
```

What names were used in the 19th century but have not been used in the 21sth century?

Produce a chart that shows the relative frequency of the names "Donald", "Hilary", "Hillary", "Joe", "Barrack", over the years 1880 through 2017.