

# Tidyverse Problem Set

*Dae Hyun Lee*

*October 4, 2019*

## Problem 1

Load the gapminder data from the gapminder package.

How many continents are included in the data set?

```
## [1] "Africa" "Americas" "Asia" "Europe" "Oceania"
```

*There are total 5 continents in the data set.*

How many countrys are included? How many countries per continent?

```
## [1] Afghanistan Albania Algeria Angola Argentina Australia  
## 142 Levels: Afghanistan Albania Algeria Angola Argentina ... Zimbabwe
```

*There are total 142 countrys are included in the data set.*

Using the gapminder data, produce a report showing the continents in the dataset, total population per continent, and GDP per capita. Be sure that the table is properly labeled and suitable for inclusion in a printed report.

Produce a well-labeled table that summarizes GDP per capita for the countries in each continent, contrasting the years 1952 and 2007.

Product a plot that summarizes the same data as the table. There should be two plots per continent. *wdwd*

Which countries in the dataset have had periods of negative population growth? *oo*

Illustrate your answer with a table or plot. *oo*

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

Illustrate your answer with a table or plot. *j*

Table 1: Total population and per capita GDP of each continent in 2002

Continent	Total Population	GDP per Capita
Africa	659081517	2071.61
Americas	739274104	16566.92
Asia	3133292191	3234.40
Europe	558142797	18420.19
Oceania	20919651	22593.03

Table 2: GDP per capita by continents in 1952 and in 2007

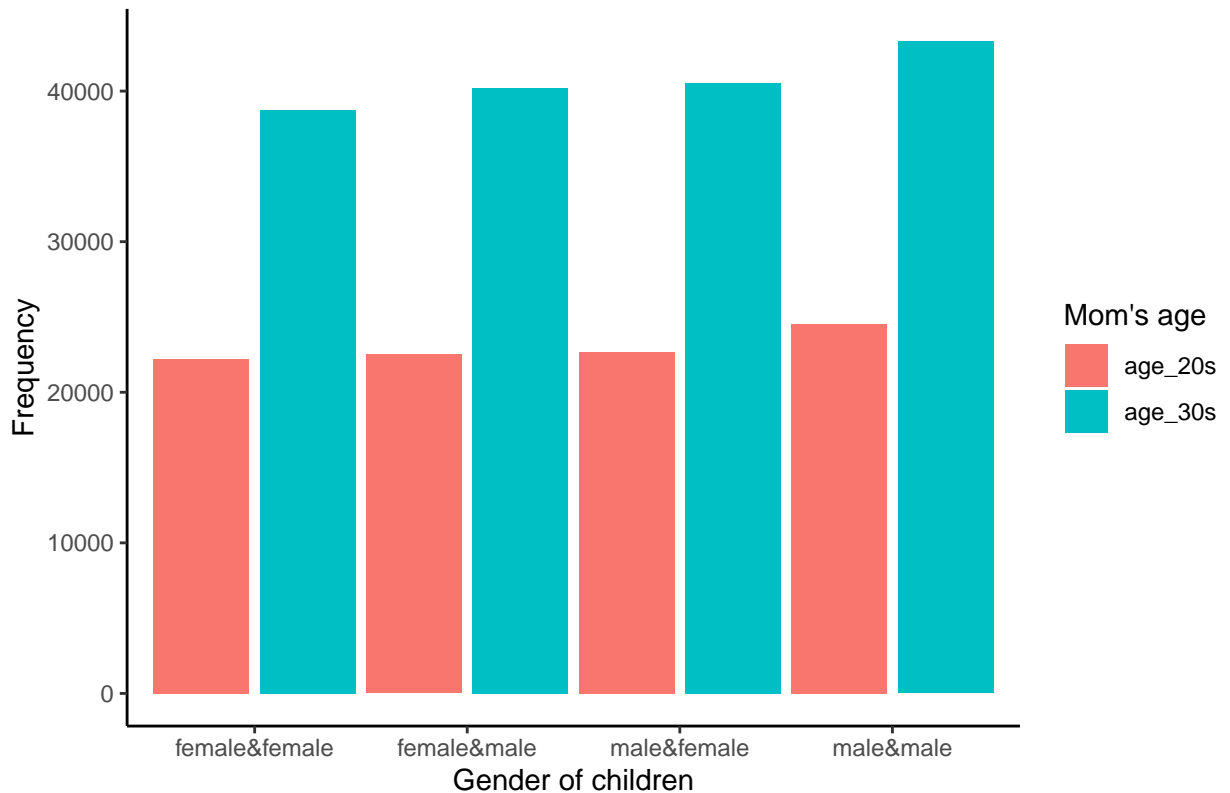
Continent	GDP per capita in 1952	GDP per capita in 2007
<b>Africa</b>	1311.22	2560.93
<b>Europe</b>	6096.66	25244.05
<b>Asia</b>	806.36	5432.37
<b>Americas</b>	8528.04	21602.75
<b>Oceania</b>	10136.10	32884.56

## 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 data 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 women who are older

### Frequency of combinations

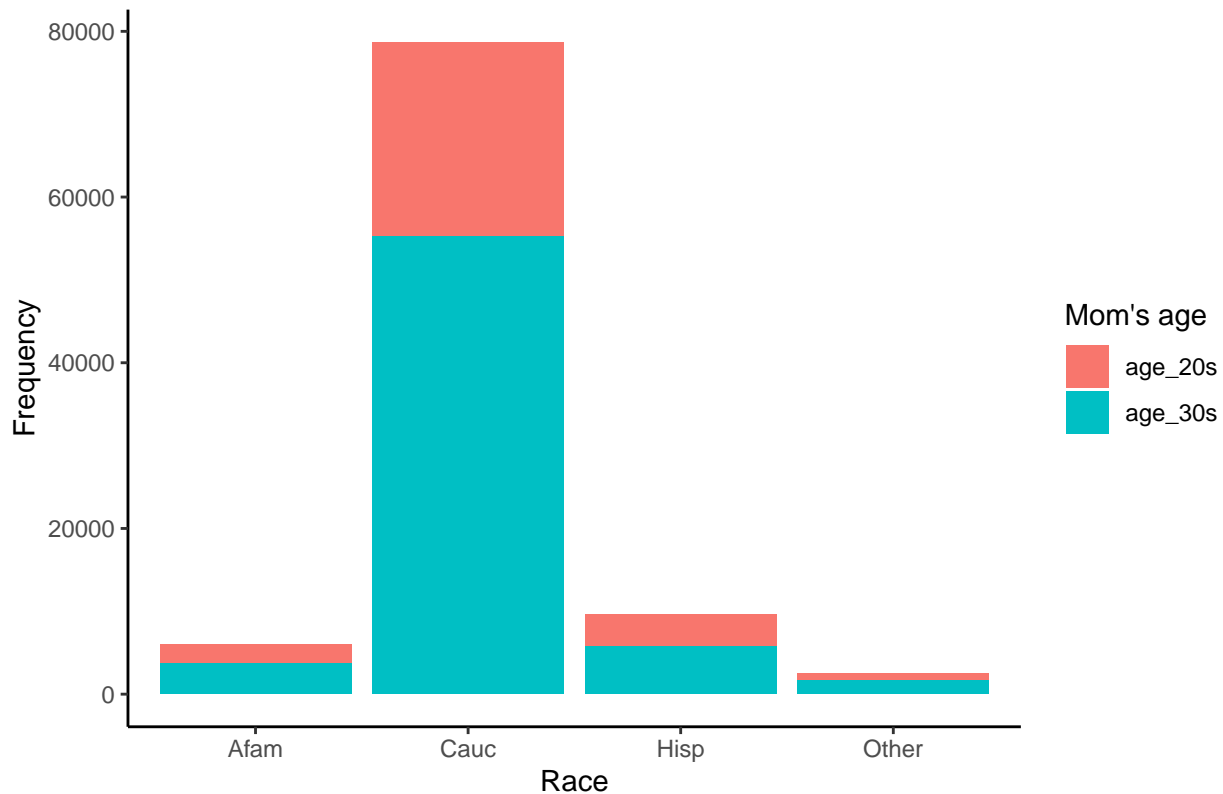


than 29?

*It is clear that there are differences in frequency between two age groups.*

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

### Frequency of having more than two children by race



### Problem 3

Use the mtcars and mpg datasets.~~ How many times does the letter “e” occur in mtcars rownames?

```
data(mtcars)
data(mpg)
```

How many cars in mtcars have the brand Merc?

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

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

### Problem 4

Install the babynames package.

```
library(babynames)
data(babynames)
babyn <- as_tibble(babynames)
babyn
```

```
## # A tibble: 1,924,665 x 5
##   year sex  name      n  prop
##   <dbl> <chr> <chr>   <int> <dbl>
## 1  1880 F    Mary    7065 0.0724
```

```
## 2 1880 F Anna 2604 0.0267
## 3 1880 F Emma 2003 0.0205
## 4 1880 F Elizabeth 1939 0.0199
## 5 1880 F Minnie 1746 0.0179
## 6 1880 F Margaret 1578 0.0162
## 7 1880 F Ida 1472 0.0151
## 8 1880 F Alice 1414 0.0145
## 9 1880 F Bertha 1320 0.0135
## 10 1880 F Sarah 1288 0.0132
## # ... with 1,924,655 more rows
```

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

```
##hist(babyn$year)
s <- sample(x = 1:1924665, size = 500000, replace = FALSE)
babyn_500000 <- babyn %>% slice(s)
babyn_500000
```

```
## # A tibble: 500,000 x 5
##   year sex name n prop
##   <dbl> <chr> <chr> <int> <dbl>
## 1 2011 F Shaleigh 6 0.0000031
## 2 1974 M Myles 75 0.0000460
## 3 1926 F Dortha 263 0.000214
## 4 1936 M Titus 18 0.0000169
## 5 1917 F Suzette 10 0.0000089
## 6 1967 F Kolette 10 0.00000583
## 7 1998 M Mckinley 90 0.0000444
## 8 2002 F Corneisha 8 0.00000405
## 9 1928 M Dominick 474 0.000415
## 10 1946 F Una 49 0.0000304
## # ... with 499,990 more rows
```

Produce a tabble that displays the five most popular boy names and girl names in the years 1880,1920, 1960, 2000.

What names overlap boys and girls?

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

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