

# 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

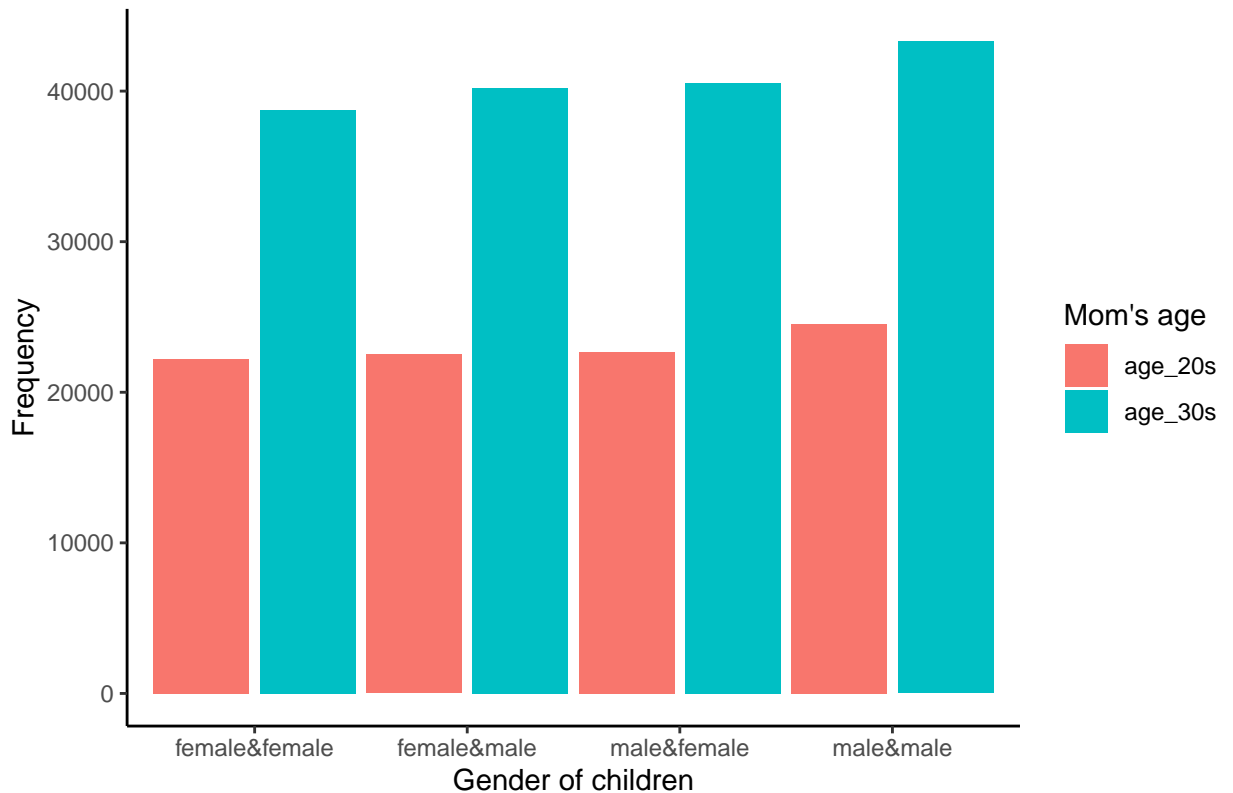
| Continent       | GDP per capita in 1952 | GDP per capita in 2007 |
|-----------------|------------------------|------------------------|
| <b>Africa</b>   | 1311.22                | 2560.93                |
| <b>Europe</b>   | 6096.66                | 25244.05               |
| <b>Asia</b>     | 806.36                 | 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 that contrasts the frequency of these four combinations. Are the frequencies different for women in their 20s and women who are older

### Frequency of combinations

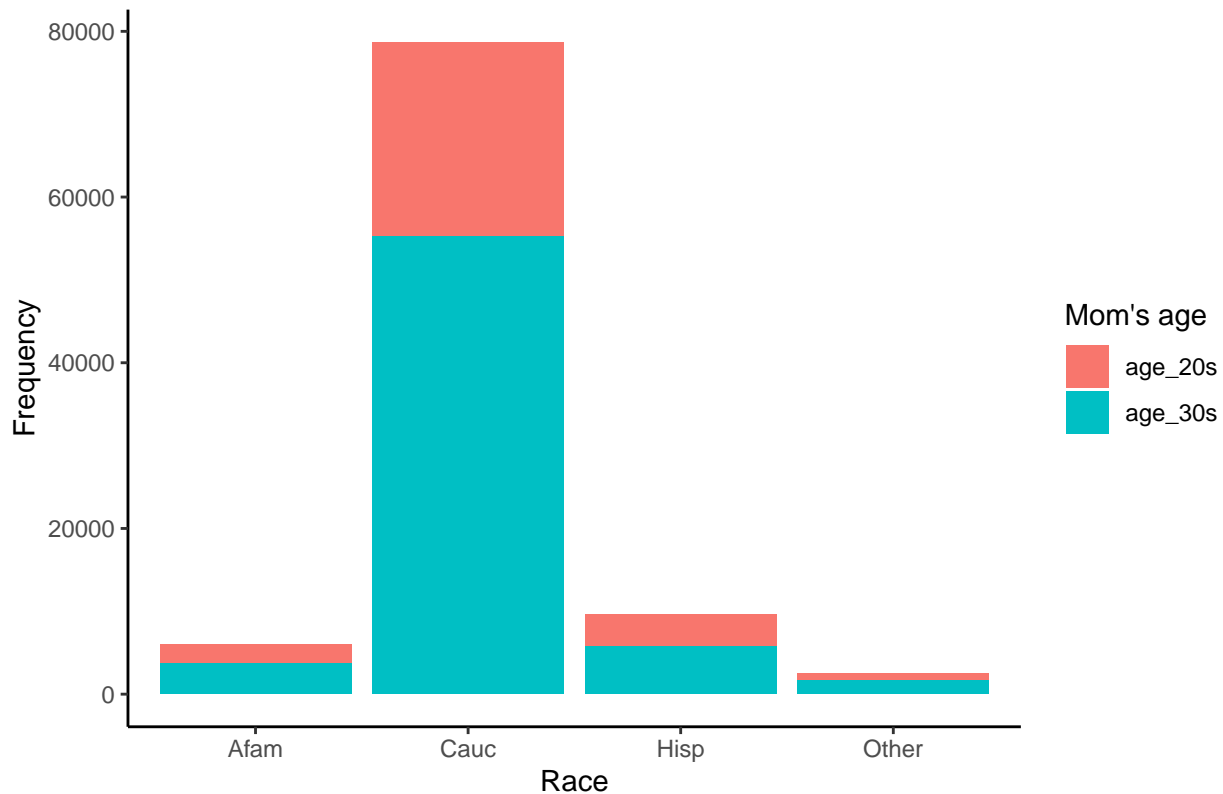


than 29?

*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?

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

*The letter “e” occurs 25 times in mtcars rownames*

How many cars in mtcars have the brand Merc?

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

*7 cars in mtcars have the brand Merc*

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

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

*There are no such cars that uses the brand Merc in mpg. Instead, there are 4 cars that have brand name mercury*

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.

Table 3: Mileage data for Merc in mpg

| Manufacturer | Model           | City miles per gallon | Highways miles per gallon |
|--------------|-----------------|-----------------------|---------------------------|
| mercury      | mountaineer 4wd | 14                    | 17                        |
| mercury      | mountaineer 4wd | 13                    | 19                        |
| mercury      | mountaineer 4wd | 13                    | 19                        |
| mercury      | mountaineer 4wd | 13                    | 17                        |

Table 4: Mileage data for Merc in mtcars

| Model       | Miles/(US)gallon |
|-------------|------------------|
| 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             |

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

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 2002 M    Minhquan    6 0.0000029
## 2 1946 M     Jame    39 0.0000236
## 3 1935 F     Candy    11 0.0000101
## 4 1923 F   Dorothe    36 0.0000287
## 5 2014 F     Mazie   127 0.0000651
## 6 1942 M   Castulo     6 0.00000426
## 7 2003 F    Terica    16 0.00000798
## 8 1933 M    Armon    17 0.0000167
## 9 1976 F    Sakia     9 0.00000573
## 10 1933 F   Deetta    14 0.0000134
## # ... with 499,990 more rows
```

Produce a tibble 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.