# Tidyverse_Yuanyuan Lin

*Yuanyuan Lin*

*2019/10/3*

## Question 1

(1).There are five continents in the data set

```
data<-gapminder
unique(data$continent)
```

```
## [1] Asia     Europe   Africa   Americas Oceania
## Levels: Africa Americas Asia Europe Oceania
```

(2).There are 142 countries included in the data set.

```
unique1<-unique(data$country)
unique1
```

```
##   [1] Afghanistan              Albania
##   [3] Algeria                  Angola
##   [5] Argentina                Australia
##   [7] Austria                  Bahrain
##   [9] Bangladesh               Belgium
##  [11] Benin                    Bolivia
##  [13] Bosnia and Herzegovina   Botswana
##  [15] Brazil                   Bulgaria
##  [17] Burkina Faso             Burundi
##  [19] Cambodia                 Cameroon
##  [21] Canada                   Central African Republic
##  [23] Chad                     Chile
##  [25] China                    Colombia
##  [27] Comoros                  Congo, Dem. Rep.
##  [29] Congo, Rep.              Costa Rica
##  [31] Cote d'Ivoire            Croatia
##  [33] Cuba                     Czech Republic
##  [35] Denmark                  Djibouti
##  [37] Dominican Republic       Ecuador
##  [39] Egypt                    El Salvador
##  [41] Equatorial Guinea        Eritrea
##  [43] Ethiopia                 Finland
##  [45] France                   Gabon
##  [47] Gambia                   Germany
##  [49] Ghana                    Greece
##  [51] Guatemala                Guinea
##  [53] Guinea-Bissau            Haiti
##  [55] Honduras                 Hong Kong, China
##  [57] Hungary                  Iceland
##  [59] India                    Indonesia
##  [61] Iran                     Iraq
##  [63] Ireland                  Israel
##  [65] Italy                    Jamaica
##  [67] Japan                    Jordan
##  [69] Kenya                    Korea, Dem. Rep.
##  [71] Korea, Rep.              Kuwait
##  [73] Lebanon                  Lesotho
##  [75] Liberia                  Libya
##  [77] Madagascar               Malawi
##  [79] Malaysia                 Mali
##  [81] Mauritania               Mauritius
##  [83] Mexico                   Mongolia
##  [85] Montenegro               Morocco
##  [87] Mozambique               Myanmar
##  [89] Namibia                  Nepal
##  [91] Netherlands              New Zealand
##  [93] Nicaragua                Niger
##  [95] Nigeria                  Norway
##  [97] Oman                     Pakistan
##  [99] Panama                   Paraguay
## [101] Peru                     Philippines
## [103] Poland                   Portugal
## [105] Puerto Rico              Reunion
## [107] Romania                  Rwanda
## [109] Sao Tome and Principe    Saudi Arabia
## [111] Senegal                  Serbia
## [113] Sierra Leone             Singapore
## [115] Slovak Republic          Slovenia
## [117] Somalia                  South Africa
## [119] Spain                    Sri Lanka
## [121] Sudan                    Swaziland
## [123] Sweden                   Switzerland
## [125] Syria                    Taiwan
## [127] Tanzania                 Thailand
## [129] Togo                     Trinidad and Tobago
## [131] Tunisia                  Turkey
## [133] Uganda                   United Kingdom
## [135] United States            Uruguay
## [137] Venezuela                Vietnam
## [139] West Bank and Gaza       Yemen, Rep.
## [141] Zambia                   Zimbabwe
## 142 Levels: Afghanistan Albania Algeria Angola Argentina ... Zimbabwe
```

(3).Countries per continent is shown in the table below

```
data%>%group_by(data$continent) %>% summarise(number = n())
```

```
## # A tibble: 5 x 2
##   `data$continent` number
##   <fct>             <int>
## 1 Africa              624
## 2 Americas            300
## 3 Asia                396
## 4 Europe              360
## 5 Oceania              24
```

(4).total population per continent and GDP per capita group by continent

```
table0<-data%>%group_by(continent)%>%summarise(mean_GPD_per_capita=mean(gdpPercap),mean_pop=mean(pop))
table0
```

```
## # A tibble: 5 x 3
##   continent mean_GPD_per_capita  mean_pop
##   <fct>                   <dbl>     <dbl>
## 1 Africa                  2194.  9916003.
## 2 Americas                7136. 24504795.
## 3 Asia                    7902. 77038722.
## 4 Europe                 14469. 17169765.
## 5 Oceania                18622.  8874672.
```

(5)GDP per capita for the countries in each continent, contrasting the years 1952 and 2007.

```
table1<-gapminder2007 <- filter(data,year == 2007)%>%group_by(continent)%>%summarise(mean_GPD_per_capita
table1
```

```
## # A tibble: 5 x 2
##   continent mean_GPD_per_capita_2007
##   <fct>                        <dbl>
## 1 Africa                       3089.
## 2 Americas                    11003.
## 3 Asia                        12473.
## 4 Europe                      25054.
## 5 Oceania                     29810.
```

```
## # A tibble: 5 x 2
##   continent mean_GPD_per_capita_1952
##   <fct>                        <dbl>
## 1 Africa                       1253.
## 2 Americas                     4079.
## 3 Asia                         5195.
## 4 Europe                       5661.
## 5 Oceania                     10298.
```

```
##   continent mean_GPD_per_capita_2007 mean_GPD_per_capita_1952
## 1    Africa                 3089.033                 1252.572
## 2  Americas                11003.032                 4079.063
## 3      Asia                12473.027                 5195.484
## 4    Europe                25054.482                 5661.057
## 5   Oceania                29810.188                10298.086
```

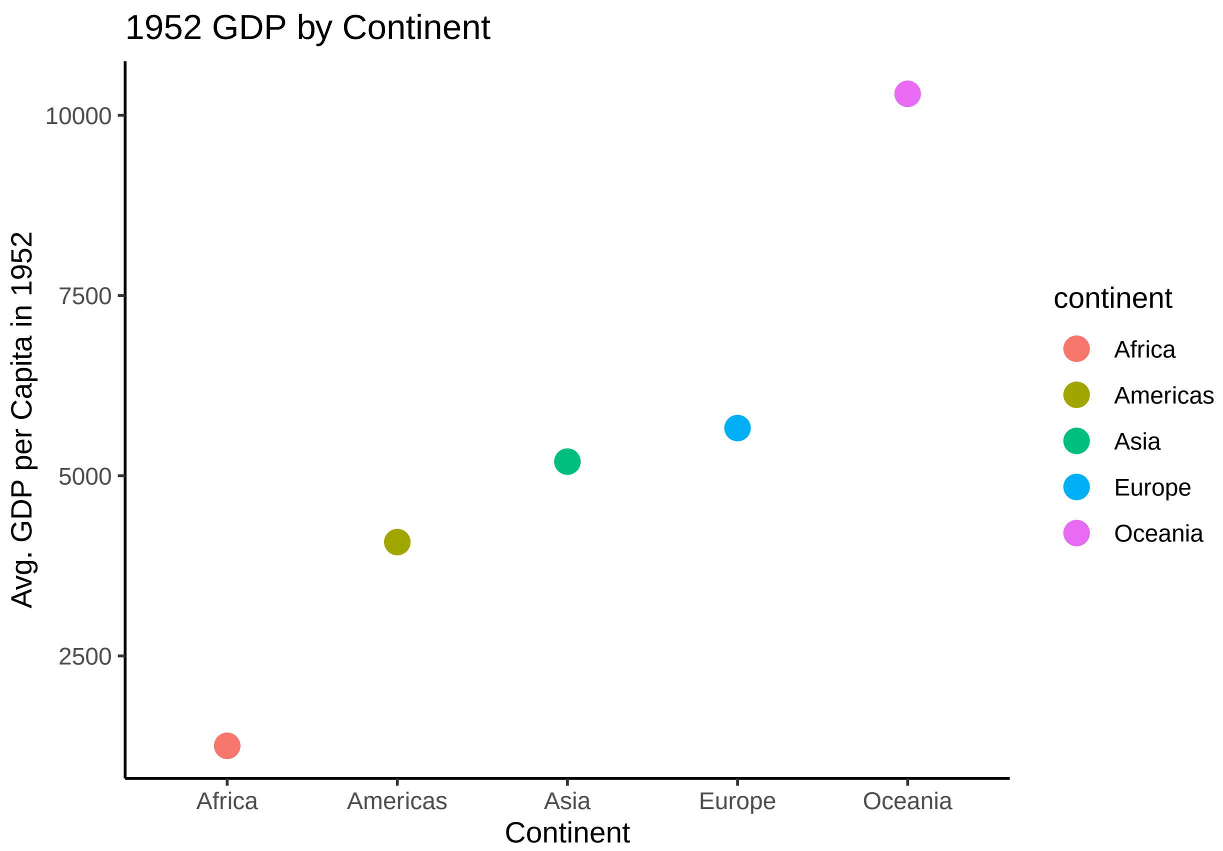
```
kable_1<-kable(new_table, format = "latex", booktabs=TRUE, digits = 2,     ## call kable to make the ta
      col.names = c("continent", "mean_GPD_per_capita_2007", "mean_GPD_per_capita_1952"),
      caption = "Total population and GDP per capita by continent" )
kable_1
```

Table 1: Total population and GDP per capita by continent

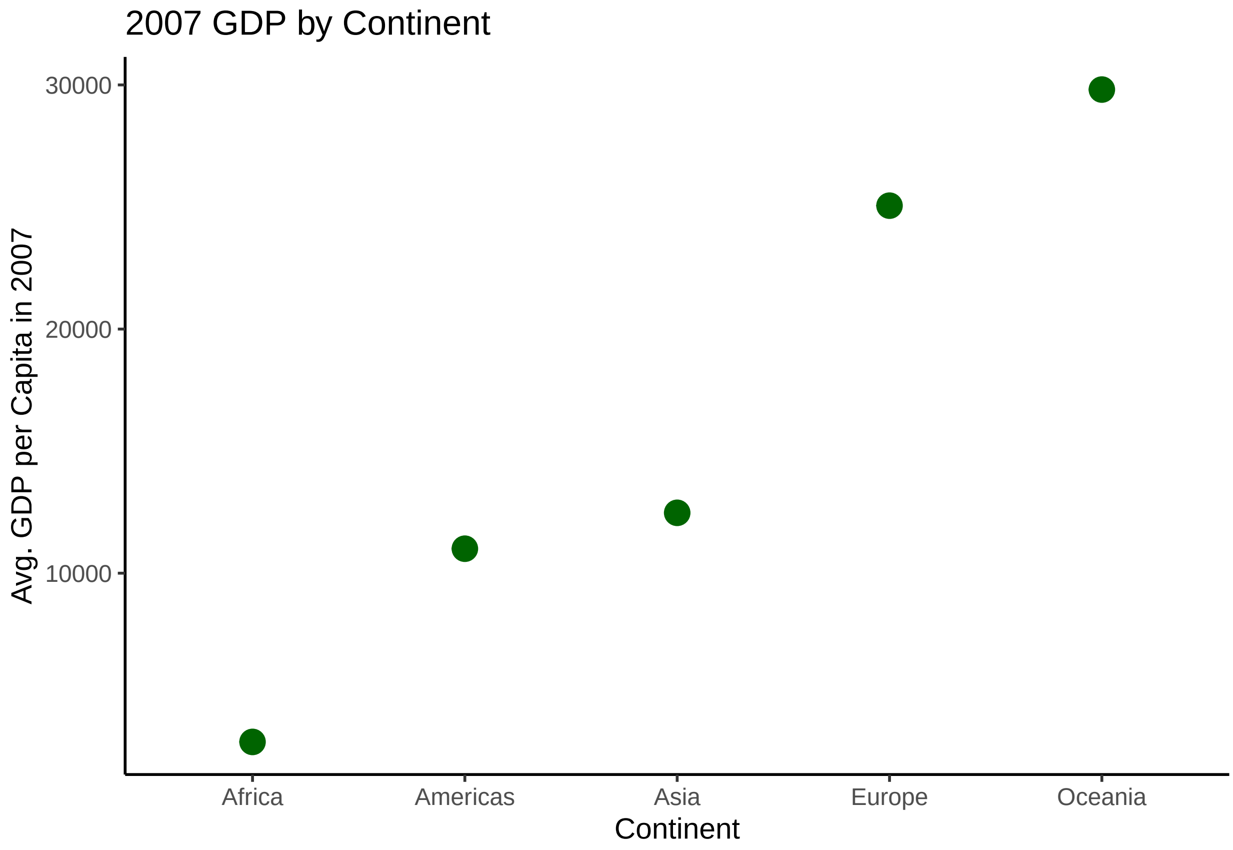
| continent | mean_GPD_per_capita_2007 | mean_GPD_per_capita_1952 |
|---|---:|---:|
| Africa | 3089.03 | 1252.57 |
| Americas | 11003.03 | 4079.06 |
| Asia | 12473.03 | 5195.48 |
| Europe | 25054.48 | 5661.06 |
| Oceania | 29810.19 | 10298.09 |

(6)plot that summarizes GDP per capita for the countries in each continent, contrasting the years 1952 and 2007.

```
ggplot(table2,aes(x = continent,y=mean_GPD_per_capita_1952,color = continent)) +
  geom_point(size=4) +
  ggtitle("1952 GDP by Continent") +
  xlab("Continent") + ylab("Avg. GDP per Capita in 1952") +
  theme_classic()
```



```
ggplot(table1,aes(x = continent,y=mean_GPD_per_capita_2007)) +
  geom_point(size=4,color="darkgreen") +
  ggtitle("2007 GDP by Continent") +
  xlab("Continent") + ylab("Avg. GDP per Capita in 2007") +
  theme_classic()
```



```
unique1<-unique(data$country)
country<-double(length(unique1))
summary<-data.frame(matrix(double(142*3),nrow = 142))
colnames(summary)<-c("Country","GDP_change","Population_change")
```

```
##                      Country  GDP_change Population_change
## 1                Afghanistan  0.25035114       2.785004462
## 2                    Albania  2.70819573       1.806994169
## 3                    Algeria  1.54117871       2.592125243
## 4                     Angola  0.36261355       1.934829204
## 5                  Argentina  1.16185054       1.254406567
## 6                  Australia  2.42995562       1.351130774
## 7                    Austria  4.88659645       0.183610402
## 8                    Bahrain  2.01974180       4.882861342
## 9                 Bangladesh  1.03327094       2.208752777
## 10                   Belgium  3.03837715       0.190348672
## 11                     Benin  0.35618150       3.647209510
## 12                   Bolivia  0.42759477       2.162731786
## 13    Bosnia and Herzegovina  6.64873642       0.631027589
## 14                  Botswana 13.76649937       2.705858813
## 15                    Brazil  3.29873875       2.356926736
## 16                  Bulgaria  3.36969732       0.006592256
## 17              Burkina Faso  1.24026001       2.204982171
## 18                   Burundi  0.26753664       2.430832207
## 19                  Cambodia  3.65107610       2.010726834
## 20                  Cameroon  0.74141005       2.532852126
## 21                   Canada  2.19510163       1.258290305
## 22 Central African Republic -0.34097874       2.382406838
## 23                     Chad  0.44575633       2.816943912
## 24                    Chile  2.34307354       1.553420171
## 25                    China 11.38389825       1.370608591
## 26                 Colombia  2.26781917       2.580954582
## 27                  Comoros -0.10593293       3.618542771
## 28         Congo, Dem. Rep. -0.64441152       3.582038021
## 29              Congo, Rep.  0.70893922       3.445755862
## 30               Costa Rica  2.67149853       3.462709850
## 31            Cote d'Ivoire  0.11245569       5.050820972
## 32                  Croatia  3.68679519       0.157405192
## 33                     Cuba  0.60172573       0.900361647
## 34           Czech Republic  2.32065776       0.120935766
## 35                  Denmark  2.63980773       0.261679742
## 36                 Djibouti -0.21990688       6.860362001
## 37       Dominican Republic  3.31086848       2.740797946
## 38                  Ecuador  0.95146118       2.876201020
## 39                    Egypt  2.93367121       2.611727803
## 40              El Salvador  0.87919433       2.397037004
## 41        Equatorial Guinea 31.35541662       1.540518243
## 42                  Eritrea  0.94980373       2.410287331
## 43                 Ethiopia  0.90753189       2.667710244
## 44                  Finland  4.16880470       0.280640508
## 45                   France  3.33440159       0.438633892
## 46                    Gabon  2.07594198       2.458188932
## 47                   Gambia  0.55132350       4.938235087
## 48                  Germany  3.50305981       0.191696601
## 49                    Ghana  0.45683140       3.098429296
## 50                   Greece  6.79972508       0.384448970
## 51                Guatemala  1.13572578       2.995996670
## 52                   Guinea  0.84762974       2.733815420
## 53            Guinea-Bissau  0.93173629       1.535147498
## 54                    Haiti -0.34706654       1.655894384
## 55                 Honduras  0.61660599       3.931792286
## 56         Hong Kong, China 12.00573037       2.283509102
## 57                  Hungary  2.42136406       0.047570286
## 58                  Iceland  3.97830769       1.040598262
## 59                    India  3.48657899       1.984936374
## 60                Indonesia  3.72287343       1.724455224
## 61                     Iran  2.82354794       3.021165470
## 62                     Iraq  0.08264290       4.053440005
## 63                  Ireland  6.80687291       0.391893247
## 64                   Israel  5.24572101       2.964848845
## 65                    Italy  4.79342492       0.219899572
## 66                  Jamaica  1.52572098       0.949471809
## 67                    Japan  8.84037850       0.474316556
## 68                   Jordan  1.92160990       8.957317976
## 69                    Kenya  0.71432822       4.508960951
## 70         Korea, Dem. Rep.  0.46384089       1.628363492
## 71              Korea, Rep. 21.65507069       1.341311553
## 72                   Kuwait -0.56351760      14.659743750
## 73                  Lebanon  1.16369858       1.724000697
## 74                  Lesotho  4.25130110       1.688022790
## 75                   Liberia -0.27983532       2.699655279
## 76                     Libya  4.05015982       4.920116031
## 77                Madagascar -0.27597946       3.024356108
## 78                    Malawi  1.05693862       3.567506294
## 79                  Malaysia  5.79997385       2.678111392
## 80                      Mali  1.30487800       2.134775497
## 81                Mauritania  1.42647408       2.197932436
## 82                 Mauritius  4.56770210       1.421580622
## 83                    Mexico  2.44368680       2.606016053
## 84                  Mongolia  2.93580309       2.589683800
## 85                Montenegro  2.49522074       0.654615136
## 86                   Morocco  1.26286409       2.396361605
## 87                Mozambique  0.75803595       2.095047776
## 88                   Myanmar  1.85196375       1.377046211
## 89                   Namibia  0.98494069       3.230030607
## 90                     Nepal  0.99931912       2.147473639
## 91               Netherlands  3.11537635       0.596092482
## 92               New Zealand  1.38571767       1.063256156
## 93                 Nicaragua -0.11664541       3.868248999
## 94                     Niger -0.18664698       2.815649386
## 95                   Nigeria  0.86949896       3.077139183
## 96                    Norway  3.88906670       0.390716429
## 97                      Oman 11.20644510       5.310927017
## 98                  Pakistan  2.80654171       3.093946800
## 99                    Panama  2.95471029       2.448826696
## 100                 Paraguay  1.13738660       3.285140333
## 101                     Peru  0.97122771       2.572866790
## 102              Philippines  1.50650377       3.058939401
## 103                   Poland  2.81947516       0.496984693
## 104                 Portugal  5.68432519       0.248272764
## 105              Puerto Rico  5.27156432       0.770314773
## 106                  Reunion  1.82105412       2.096988747
## 107                  Romania  2.43713995       0.339510283
## 108                   Rwanda  0.74953718       2.495401643
## 109    Sao Tome and Principe  0.81726344       2.325706954
## 110             Saudi Arabia  2.35237219       5.890480186
## 111                  Senegal  0.18072458       3.451858750
## 112                   Serbia  1.73255494       0.479598761
## 113             Sierra Leone -0.01960357       1.866937999
## 114                Singapore 19.36300861       3.039937001
## 115          Slovak Republic  2.68070327       0.530998385
## 116                 Slovenia  5.11340508       0.348922940
## 117                  Somalia -0.18455541       2.608545568
## 118             South Africa  0.96170964       2.084334278
## 119                    Spain  6.51716289       0.416755698
## 120                Sri Lanka  2.66403142       1.552914796
## 121                    Sudan  0.61040178       3.972908287
## 122                Swaziland  2.93031392       2.903852978
## 123                   Sweden  2.97049310       0.267579298
## 124              Switzerland  1.54552916       0.568984631
## 125                    Syria  1.54614261       4.275020763
## 126                   Taiwan 22.79413107       1.710328990
## 127                 Tanzania  0.54535976       3.582480318
## 128                 Thailand  8.84220341       2.056363396
## 129                     Togo  0.02693772       3.676825692
## 130      Trinidad and Tobago  4.95662900       0.594037867
## 131                  Tunisia  3.83012648       1.817133920
## 132                   Turkey  3.29550159       2.200201505
## 133                   Uganda  0.43773408       4.007968175
## 134           United Kingdom  2.32714395       0.205160381
## 135            United States  2.07006241       0.911356477
## 136                  Uruguay  0.85620010       0.530203976
## 137                Venezuela  0.48453875       3.795355440
## 138                  Vietnam  3.03521999       2.248480931
## 139       West Bank and Gaza  0.99615011       2.899078679
## 140              Yemen, Rep.  1.91763928       3.474719617
## 141                   Zambia  0.10791700       3.395971183
## 142                 Zimbabwe  0.15440559       2.995947622
```

```
negative_gro <- filter(summary1,Population_change<0)
negative_gro
```

```
## [1] Country           GDP_change        Population_change
## <0 rows> (or 0-length row.names)
```

```
max_gdp<-filter(summary,GDP_change==max(GDP_change))
max_gdp
```

```
##             Country GDP_change Population_change
## 1 Equatorial Guinea   31.35542          1.540518
```

```
nrow(data)
```

```
## [1] 1704
```

```
for(i in 1:nrow(data)){
  a=data$pop[i]
  b=data$pop[i+1]
  c <- (b-a)/a
  data$negative_check[i]<-c
}
```

```
## Warning: Unknown or uninitialised column: 'negative_check'.
```

```
country.unique<-unique(data$country)
for(p in 1:length(country.unique)){
  if(data$country[p]!=data$country[p+1]){
    data$negative_check[p]<-1
  }
}
```

```
getcountry<-rep(0,1703)
for(g in 1:length(getcountry)){
  if(data$negative_check[g]<0){
    getcountry<-data$country[g]
  }
}
getcountry
```

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

```
getyear<-rep(0,1703)
for(g in 1:length(getyear)){
  if(data$negative_check[g]<0){
    getyear<-data$year[g]
  }
}
getyear
```

```
## [1] 2007
```

```
for(i in 1:1703){
  e=data$gdpPercap[i]
  f=data$gdpPercap[i+1]
  g <- (f-e)/e
  data$gdp_check[i]<-g
}
```

```
## Warning: Unknown or uninitialised column: 'gdp_check'.
```

```
for(p in 1:length(country.unique)){
  if(data$country[p]!=data$country[p+1]){
    data$gdp_check[p]<-0
  }
}
```

```
max(data$gdp_check,na.rm=TRUE)
```

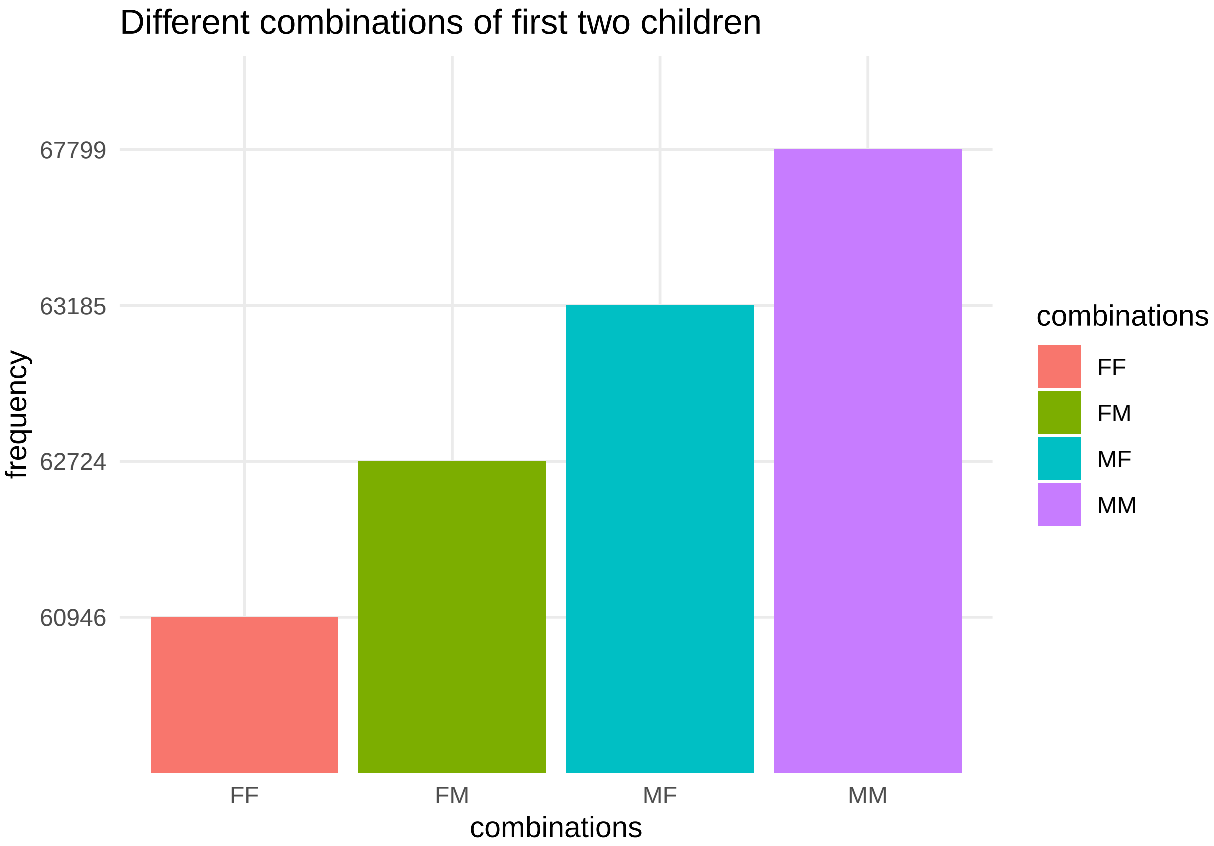
```
## [1] 8.49069
```

```
#mac_gdp_cou<-filter(data,gdp_check==8.49069)
#mac_gdp_cou
#answer is Gambia, Africa
```

## Question 2

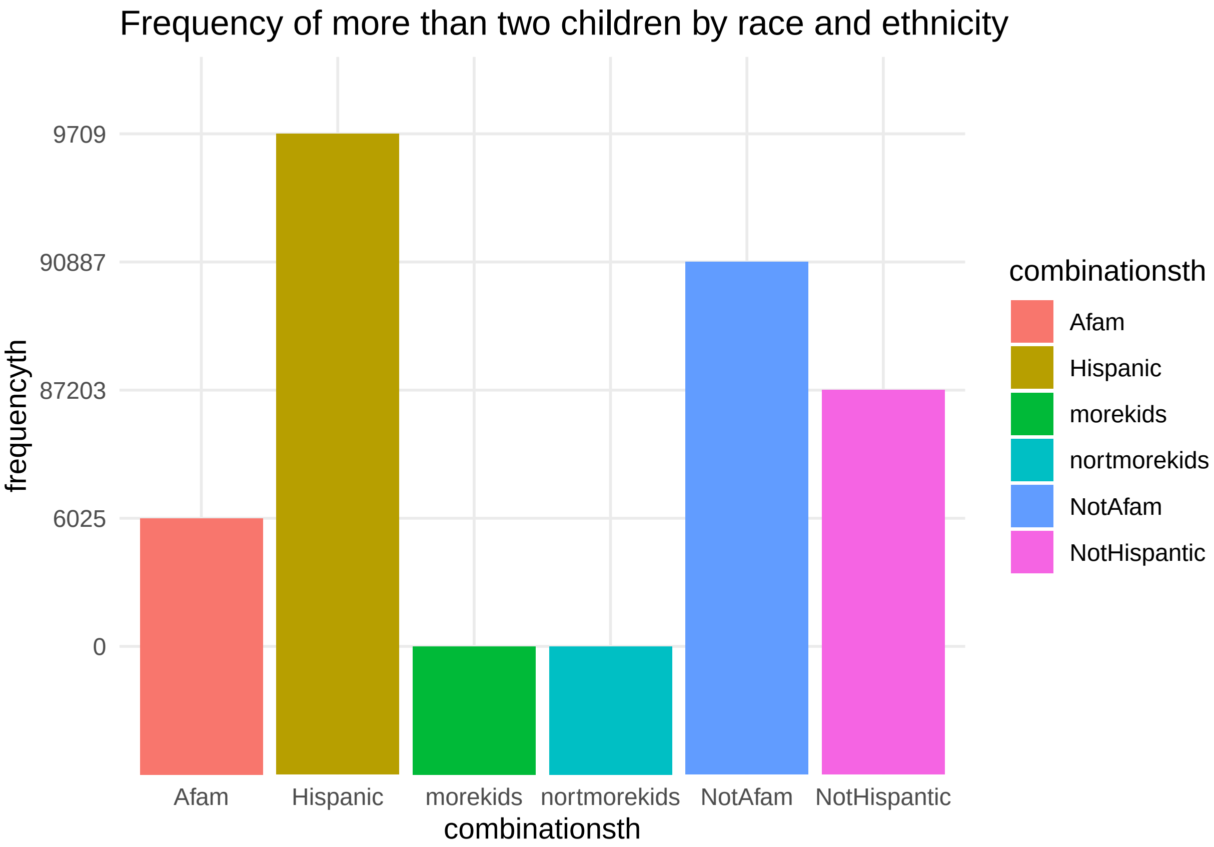
```
#d<-data("GSS7402", package = "AER")
data('Fertility')
data2<-Fertility
MM<-data2[data2$gender1=='male' & data2$gender2=='male',]
MF<-data2[data2$gender1=='male' & data2$gender2=='female',]
```

```r
FF<-data2[data2$gender1=='female' & data2$gender2=='female',]
FM<-data2[data2$gender1=='female' & data2$gender2=='male',]
frequency<-c(nrow(MM),nrow(MF),nrow(FF),nrow(FM))
combinations<-c("MM","MF","FF","FM")
da<-data.frame(cbind(frequency,combinations))
ggplot(da,aes(y=frequency,x=combinations,fill=combinations))+
  geom_bar(stat="identity")+theme_minimal()+
  ggtitle("Different combinations of first two children")
```



```r
Hispanic<-data2[data2$morekids=='yes' & data2$hispanic=='yes',]
NotHispanic<-data2[data2$morekids=='yes' & data2$hispanic=='no',]
Afam<-data2[data2$morekids=='yes' & data2$afam =='yes',]
NotAfam<-data2[data2$morekids=='yes' & data2$afam =='no',]
notmorekids<-data2[data2$morekids=='yes' & data2$work =='yes',]
morekids<-data2[data2$morekids=='yes' & data2$work =='no',]


frequencyth<-c(nrow(Hispanic),nrow(NotHispanic),nrow(Afam),nrow(NotAfam),nrow(notmorekids),nrow(morekids
combinationsth<-c("Hispanic","NotHispantic","Afam","NotAfam","nortmorekids","morekids")
dat<-data.frame(cbind(frequencyth,combinationsth))
ggplot(dat,aes(y=frequencyth,x=combinationsth,fill=combinationsth))+
  geom_bar(stat="identity")+theme_minimal()+
  ggtitle("Frequency of more than two children by race and ethnicity")
```



```
twoMM<-data2[data2$gender1=='male' & data2$gender2=='male'& data2$age<29,]
twoMF<-data2[data2$gender1=='male' & data2$gender2=='female'& data2$age<29,]
twoFF<-data2[data2$gender1=='female' & data2$gender2=='female'& data2$age<29,]
twoFM<-data2[data2$gender1=='female' & data2$gender2=='male'& data2$age<29,]
frequency1<-c(nrow(twoMM),nrow(twoMF),nrow(twoFF),nrow(twoFM))
combinations1<-c("twoMM","twoMF","twoFF","twoFM")
da<-data.frame(cbind(frequency1,combinations1))

thirMM<-data2[data2$gender1=='male' & data2$gender2=='male'& data2$age>29,]
thirMF<-data2[data2$gender1=='male' & data2$gender2=='female'& data2$age>29,]
thirFF<-data2[data2$gender1=='female' & data2$gender2=='female'& data2$age>29,]
thirFM<-data2[data2$gender1=='female' & data2$gender2=='male'& data2$age>29,]
nrow(twoMM)==nrow(thirMM)
```

```
## [1] FALSE
```

```
nrow(twoMF)==nrow(thirMF)
```

```
## [1] FALSE
```

```
nrow(twoFF)==nrow(thirFF)
```

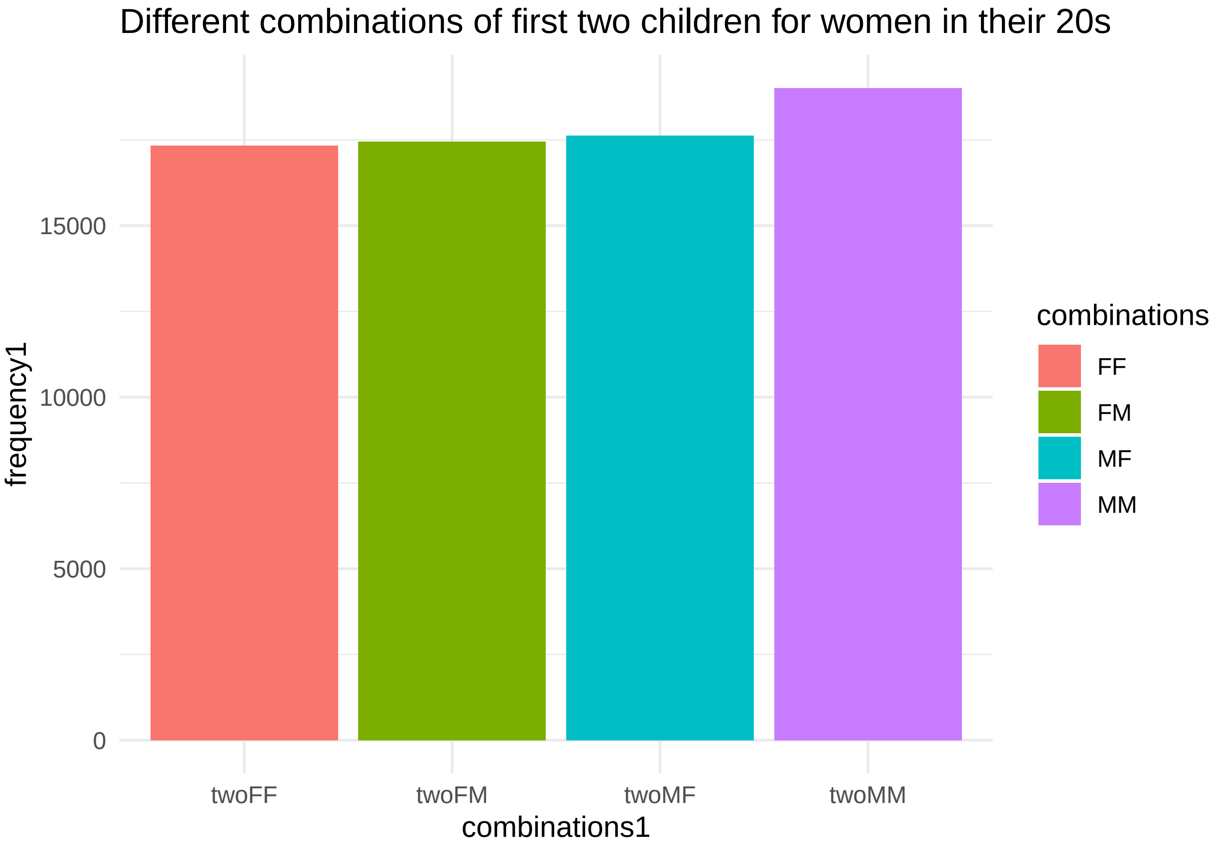
```
## [1] FALSE
```

```
nrow(twoFM)==nrow(thirFM)
```

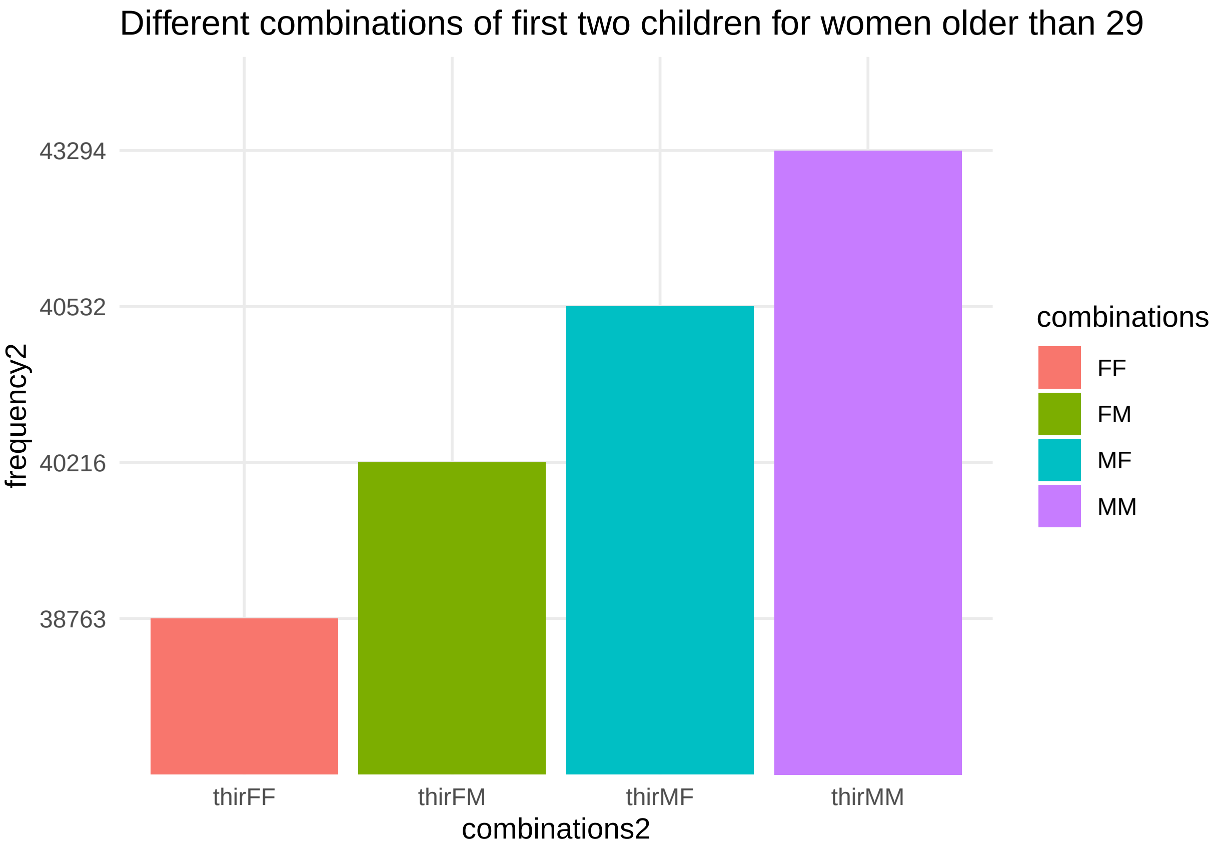
```
## [1] FALSE
```

```
frequency2<-c(nrow(thirMM),nrow(thirMF),nrow(thirFF),nrow(thirFM))
combinations2<-c("thirMM","thirMF","thirFF","thirFM")
da<-data.frame(cbind(frequency2,combinations2))

par(mfrow=c(1,2))
ggplot(da,aes(y=frequency1,x=combinations1,fill=combinations))+
  geom_bar(stat="identity")+theme_minimal()+
  ggtitle("Different combinations of first two children for women in their 20s")
```



```
ggplot(da,aes(y=frequency2,x=combinations2,fill=combinations))+
  geom_bar(stat="identity")+theme_minimal()+
  ggtitle("Different combinations of first two children for women older than 29")
```



### Question 3

```r
library(knitr)
datt<-mtcars
dattt<-mpg
carname<-rownames(datt)
library(stringr)
sum(str_count(carname,'e'))
```

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

```r
sum(str_count(carname,'Merc'))
```

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

```r
sum(str_count(dattt$manufacturer,'mercury'))
```

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

```r
#mercmtcars<-mtcars[mtcars$]

# [1]  3 4 3 3 2 3 1 1 2 2
mtcarsmerc<-datt[which(str_count(carname,'Merc') %in% c(1)),]
mtcarsmerc
```

```
##                 mpg cyl  disp  hp drat   wt qsec vs am gear carb
## Merc 240D      24.4   4 146.7  62 3.69 3.19 20.0  1  0    4    2
## Merc 230       22.8   4 140.8  95 3.92 3.15 22.9  1  0    4    2
## Merc 280       19.2   6 167.6 123 3.92 3.44 18.3  1  0    4    4
## Merc 280C      17.8   6 167.6 123 3.92 3.44 18.9  1  0    4    4
## Merc 450SE     16.4   8 275.8 180 3.07 4.07 17.4  0  0    3    3
## Merc 450SL     17.3   8 275.8 180 3.07 3.73 17.6  0  0    3    3
## Merc 450SLC    15.2   8 275.8 180 3.07 3.78 18.0  0  0    3    3
```

```
mpgmerc<-dattt[which(str_count(dattt$manufacturer,'mercury') %in% c(1)),]
mpgmerc
```

```
## # A tibble: 4 x 11
##   manufacturer model  displ  year   cyl trans drv     cty   hwy fl    class
##   <chr>        <chr>  <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
## 1 mercury      mount~   4    1999     6 auto~ 4        14    17 r     suv
## 2 mercury      mount~   4    2008     6 auto~ 4        13    19 r     suv
## 3 mercury      mount~   4.6  2008     8 auto~ 4        13    19 r     suv
## 4 mercury      mount~   5    1999     8 auto~ 4        13    17 r     suv
```

```
l<-data.frame(unclass(summary(mtcarsmerc$mpg)), check.names = FALSE, stringsAsFactors = FALSE)
table<-knitr::kable(l)
table
```

| | unclass(summary(mtcarsmerc$mpg)) |
|---|---:|
| Min. | 15.20000 |
| 1st Qu. | 16.85000 |
| Median | 17.80000 |
| Mean | 19.01429 |
| 3rd Qu. | 21.00000 |
| Max. | 24.40000 |

```
#colnames(table)<-c('summary','v')
```

## Question 4

```
library(babynames)
baby_names<-babynames
```

```
sample<-baby_names[sample(nrow(baby_names), 500000), ]
```

five most popular boy names and girl names in year 1880

```
## # A tibble: 295 x 3
## # Groups:   sex_f [1]
##    sex_f name_f  total_f
##    <chr> <chr>     <int>
##  1 M     James      5927
##  2 M     George     5126
##  3 M     Henry      2444
##  4 M     Edward     2364
##  5 M     Harry      2152
##  6 M     Walter     1755
##  7 M     Samuel     1024
##  8 M     Louis       828
##  9 M     Joe         731
## 10 M     Charlie     730
## # ... with 285 more rows

## # A tibble: 5 x 3
## # Groups:   sex_f [1]
##   sex_f name_f total_f
##   <chr> <chr>    <int>
## 1 M     James     5927
## 2 M     George    5126
## 3 M     Henry     2444
## 4 M     Edward    2364
## 5 M     Harry     2152

## # A tibble: 262 x 3
## # Groups:   sex [1]
##    sex   name      total
##    <chr> <chr>     <int>
##  1 F     Mary       7065
##  2 F     Emma       2003
##  3 F     Elizabeth  1939
##  4 F     Clara      1226
##  5 F     Ella       1156
##  6 F     Laura      1012
##  7 F     Grace       982
##  8 F     Carrie      949
##  9 F     Julia       783
## 10 F     Hattie      769
## # ... with 252 more rows

## # A tibble: 262 x 3
## # Groups:   sex_m [1]
##    sex_m name_m    total_m
##    <chr> <chr>       <int>
##  1 F     Mary         7065
##  2 F     Emma         2003
##  3 F     Elizabeth    1939
##  4 F     Clara        1226
##  5 F     Ella         1156
##  6 F     Laura        1012
##  7 F     Grace         982
##  8 F     Carrie        949
##  9 F     Julia         783
## 10 F     Hattie        769
## # ... with 252 more rows

## # A tibble: 5 x 3
```

Table 2: Most popular boy and girl names in 1880

| sex_f | name_f | total_f | sex_m | name_m | total_m |
|---|---|---|---|---|---|
| M | James | 5927 | F | Mary | 7065 |
| M | George | 5126 | F | Emma | 2003 |
| M | Henry | 2444 | F | Elizabeth | 1939 |
| M | Edward | 2364 | F | Clara | 1226 |
| M | Harry | 2152 | F | Ella | 1156 |

```
## # Groups:   sex_m [1]
##   sex_m name_m    total_m
##   <chr> <chr>       <int>
## 1 F     Mary         7065
## 2 F     Emma         2003
## 3 F     Elizabeth    1939
## 4 F     Clara        1226
## 5 F     Ella         1156
```

```
new_table1<-data.frame(cbind(top5_1880_female,top5_1880_male))
new_table1
```

```
##   sex_f name_f total_f sex_m    name_m total_m
## 1     M  James    5927     F      Mary    7065
## 2     M George    5126     F      Emma    2003
## 3     M  Henry    2444     F Elizabeth    1939
## 4     M Edward    2364     F     Clara    1226
## 5     M  Harry    2152     F      Ella    1156
```

```
kable_2<-kable(new_table1, format = "latex", booktabs=TRUE, digits = 2,     ## call kable to make the ta
      col.names = c("sex_f", "name_f", "total_f","sex_m","name_m","total_m"),
      caption = "Most popular boy and girl names in 1880" )
kable_2
```

five most popular boy names and girl names in year 1920

```
## # A tibble: 1,279 x 3
## # Groups:   sex_f [1]
##    sex_f name_f  total_f
##    <chr> <chr>     <int>
##  1 M     Harold    13679
##  2 M     Paul      12569
##  3 M     Raymond   12194
##  4 M     Arthur    10236
##  5 M     Harry      9408
##  6 M     Earl       6532
##  7 M     Roy        6353
##  8 M     Francis    6241
##  9 M     Joe        6071
## 10 M     Leonard    5511
## # ... with 1,269 more rows
## # A tibble: 5 x 3
## # Groups:   sex_f [1]
##   sex_f name_f  total_f
##   <chr> <chr>     <int>
## 1 M     Harold    13679
## 2 M     Paul      12569
## 3 M     Raymond   12194
## 4 M     Arthur    10236
## 5 M     Harry      9408
```

```
## # A tibble: 1,475 x 3
## # Groups:   sex [1]
##    sex   name      total
##    <chr> <chr>     <int>
##  1 F     Mary      70980
##  2 F     Dorothy   36643
##  3 F     Margaret  27997
##  4 F     Ruth      26101
##  5 F     Elizabeth 15910
##  6 F     Frances   15883
##  7 F     Anna      14580
##  8 F     Evelyn    13838
##  9 F     Marjorie   8659
## 10 F     Eleanor    8498
## # ... with 1,465 more rows
```

```
## # A tibble: 262 x 3
## # Groups:   sex_m [1]
##    sex_m name_m    total_m
##    <chr> <chr>       <int>
##  1 F     Mary         7065
##  2 F     Emma         2003
##  3 F     Elizabeth    1939
##  4 F     Clara        1226
##  5 F     Ella         1156
##  6 F     Laura        1012
##  7 F     Grace         982
##  8 F     Carrie        949
##  9 F     Julia         783
## 10 F     Hattie        769
## # ... with 252 more rows
```

```
## # A tibble: 5 x 3
## # Groups:   sex_m [1]
##   sex_m name_m    total_m
##   <chr> <chr>       <int>
## 1 F     Mary         7065
## 2 F     Emma         2003
## 3 F     Elizabeth    1939
## 4 F     Clara        1226
## 5 F     Ella         1156
```

Table 3: Most five popular boy and girl names in 1920

| sex_f | name_f | total_f | sex_m | name_m | total_m |
|---|---|---|---|---|---|
| M | Harold | 13679 | F | Mary | 7065 |
| M | Paul | 12569 | F | Emma | 2003 |
| M | Raymond | 12194 | F | Elizabeth | 1939 |
| M | Arthur | 10236 | F | Clara | 1226 |
| M | Harry | 9408 | F | Ella | 1156 |

```
new_table2<-data.frame(cbind(top5_1920_female,top5_1920_male))
new_table2
```

```
##   sex_f  name_f total_f sex_m    name_m total_m
## 1     M  Harold   13679     F      Mary    7065
## 2     M    Paul   12569     F      Emma    2003
## 3     M Raymond   12194     F Elizabeth    1939
## 4     M  Arthur   10236     F     Clara    1226
## 5     M   Harry    9408     F      Ella    1156
```

```
kable_3<-kable(new_table2, format = "latex", booktabs=TRUE, digits = 2,     ## call kable to make the ta
      col.names = c("sex_f", "name_f", "total_f","sex_m","name_m","total_m"),
      caption = "Most five popular boy and girl names in 1920" )
kable_3
```

five most popular boy names and girl names in year 1960

```
## # A tibble: 1,178 x 3
## # Groups:   sex_f [1]
##    sex_f name_f  total_f
##    <chr> <chr>     <int>
##  1 M     David     85928
##  2 M     William   49354
##  3 M     Joseph    29746
##  4 M     Ronald    21700
##  5 M     Gary      21684
##  6 M     Scott     21600
##  7 M     Gregory   20316
##  8 M     Larry     15988
##  9 M     Douglas   15147
## 10 M     Mike      12047
## # ... with 1,168 more rows
```

```
## # A tibble: 5 x 3
## # Groups:   sex_f [1]
##   sex_f name_f  total_f
##   <chr> <chr>     <int>
## 1 M     David     85928
## 2 M     William   49354
## 3 M     Joseph    29746
## 4 M     Ronald    21700
## 5 M     Gary      21684
## # A tibble: 1,881 x 3
## # Groups:   sex [1]
##    sex   name      total
##    <chr> <chr>     <int>
##  1 F     Lori      18685
##  2 F     Carol     17460
##  3 F     Denise    15065
##  4 F     Cindy     14949
##  5 F     Tammy     14310
##  6 F     Janet     14251
##  7 F     Kim       12474
##  8 F     Catherine  9536
##  9 F     Tina       9128
## 10 F     Angela     8682
## # ... with 1,871 more rows
```

```
## # A tibble: 1,881 x 3
## # Groups:   sex_m [1]
##    sex_m name_m    total_m
##    <chr> <chr>       <int>
##  1 F     Lori        18685
##  2 F     Carol       17460
##  3 F     Denise      15065
##  4 F     Cindy       14949
##  5 F     Tammy       14310
##  6 F     Janet       14251
##  7 F     Kim         12474
##  8 F     Catherine    9536
##  9 F     Tina         9128
## 10 F     Angela       8682
## # ... with 1,871 more rows
```

```
## # A tibble: 5 x 3
## # Groups:   sex_m [1]
##   sex_m name_m total_m
##   <chr> <chr>    <int>
## 1 F     Lori     18685
## 2 F     Carol    17460
## 3 F     Denise   15065
## 4 F     Cindy    14949
## 5 F     Tammy    14310
```

```r
new_table3<-data.frame(cbind(top5_1960_female,top5_1960_male))
new_table3
```

```
##   sex_f  name_f total_f sex_m name_m total_m
## 1     M   David   85928     F   Lori   18685
## 2     M William   49354     F  Carol   17460
## 3     M  Joseph   29746     F Denise   15065
## 4     M  Ronald   21700     F  Cindy   14949
## 5     M    Gary   21684     F  Tammy   14310
```

Table 4: Most five popular boy and girl names in 1960

| sex_f | name_f | total_f | sex_m | name_m | total_m |
|---|---|---|---|---|---|
| M | David | 85928 | F | Lori | 18685 |
| M | William | 49354 | F | Carol | 17460 |
| M | Joseph | 29746 | F | Denise | 15065 |
| M | Ronald | 21700 | F | Cindy | 14949 |
| M | Gary | 21684 | F | Tammy | 14310 |

```
kable_4<-kable(new_table3, format = "latex", booktabs=TRUE, digits = 2,      ## call kable to make the ta
      col.names = c("sex_f", "name_f", "total_f","sex_m","name_m","total_m"),
      caption = "Most five popular boy and girl names in 1960" )
kable_4
```

five most popular boy names and girl names in year 2000

```
sample4<-filter(sample,year==2000)%>%
  group_by(sex,name)%>%
  summarize(total=sum(n))%>%
  arrange(desc(total))
sample4_1<-filter(sample4,sex=="M")
sample4_1_1<- rename(sample4_1, sex_f = sex,name_f=name,total_f=total)
sample4_1_1
```

```
## # A tibble: 3,192 x 3
## # Groups:   sex_f [1]
##    sex_f name_f   total_f
##    <chr> <chr>      <int>
##  1 M     Joshua     27538
##  2 M     Daniel     22312
##  3 M     Tyler      21503
##  4 M     James      17981
##  5 M     Austin     15944
##  6 M     Benjamin   14840
##  7 M     Samuel     14170
##  8 M     Jose       12581
##  9 M     Eric        9156
## 10 M     Adam        8133
## # ... with 3,182 more rows
```

```
top5_2000_female<-sample4_1_1[1:5,]
top5_2000_female
```

```
## # A tibble: 5 x 3
## # Groups:   sex_f [1]
##   sex_f name_f total_f
##   <chr> <chr>    <int>
## 1 M     Joshua   27538
## 2 M     Daniel   22312
## 3 M     Tyler    21503
## 4 M     James    17981
## 5 M     Austin   15944
## # A tibble: 4,643 x 3
## # Groups:   sex [1]
##    sex   name      total
##    <chr> <chr>     <int>
##  1 F     Madison   19967
##  2 F     Alexis    17629
##  3 F     Elizabeth 15094
##  4 F     Brianna   12878
##  5 F     Victoria  10923
##  6 F     Sydney    10242
##  7 F     Jasmine    9097
##  8 F     Julia      8766
##  9 F     Kaitlyn    8758
## 10 F     Amanda     8552
## # ... with 4,633 more rows
```

```
## # A tibble: 4,643 x 3
## # Groups:   sex_m [1]
##    sex_m name_m    total_m
##    <chr> <chr>       <int>
##  1 F     Madison     19967
##  2 F     Alexis      17629
##  3 F     Elizabeth   15094
##  4 F     Brianna     12878
##  5 F     Victoria    10923
##  6 F     Sydney      10242
##  7 F     Jasmine      9097
##  8 F     Julia        8766
##  9 F     Kaitlyn      8758
## 10 F     Amanda       8552
## # ... with 4,633 more rows
```

```
## # A tibble: 5 x 3
## # Groups:   sex_m [1]
##   sex_m name_m    total_m
##   <chr> <chr>       <int>
## 1 F     Madison     19967
## 2 F     Alexis      17629
## 3 F     Elizabeth   15094
## 4 F     Brianna     12878
## 5 F     Victoria    10923
```

```
new_table4<-data.frame(cbind(top5_2000_female,top5_2000_male))
new_table4
```

```
##   sex_f name_f total_f sex_m    name_m total_m
## 1     M Joshua   27538     F   Madison   19967
## 2     M Daniel   22312     F    Alexis   17629
## 3     M  Tyler   21503     F Elizabeth   15094
## 4     M  James   17981     F   Brianna   12878
## 5     M Austin   15944     F  Victoria   10923
```

Table 5: Most five popular boy and girl names in 2000

| sex_f | name_f | total_f | sex_m | name_m | total_m |
|---|---|---|---|---|---|
| M | Joshua | 27538 | F | Madison | 19967 |
| M | Daniel | 22312 | F | Alexis | 17629 |
| M | Tyler | 21503 | F | Elizabeth | 15094 |
| M | James | 17981 | F | Brianna | 12878 |
| M | Austin | 15944 | F | Victoria | 10923 |

```r
kable_5<-kable(new_table4, format = "latex", booktabs=TRUE, digits = 2,     ## call kable to make the ta
      col.names = c("sex_f", "name_f", "total_f","sex_m","name_m","total_m"),
      caption = "Most five popular boy and girl names in 2000" )
kable_5
```

```r
boyname<-subset(baby_names,baby_names$sex=="M",select = name)

girlname<-subset(baby_names,baby_names$sex=="F",select = name)
samename<-inner_join(boyname,girlname,by="name")
sharename<-unique(samename)
sharename
```

```
## # A tibble: 10,663 x 1
##    name
##    <chr>
##  1 John
##  2 William
##  3 James
##  4 Charles
##  5 George
##  6 Frank
##  7 Joseph
##  8 Thomas
##  9 Henry
## 10 Robert
## # ... with 10,653 more rows
```

```r
#names were used in the 19th century but have not been used in the 21sth century
nineth_cen<-filter(sample,year>"2000")
nineth_cen
name_in<-unique(nineth_cen$name)
name_in
```

```r
newdata<-filter(sample,year>1880&year<2017)
newdata
```

```
## # A tibble: 490,940 x 5
##     year sex   name          n      prop
##    <dbl> <chr> <chr>     <int>     <dbl>
##  1  2012 M     Lathen       16 0.0000079
##  2  1987 M     Justyn       68 0.0000349
##  3  1917 F     Pansy        238 0.000212
##  4  2010 F     Lauralynn      6 0.00000306
##  5  1954 M     Nicky        163 0.0000788
##  6  1930 F     Orena          6 0.00000514
##  7  1994 F     Shantonia      7 0.00000359
##  8  1983 F     Tawanna      113 0.0000632
##  9  1959 F     Santa         32 0.0000154
## 10  1936 M     Hilmar         5 0.0000047
## # ... with 490,930 more rows
```

```
Donald <- sample[sample$name=="Donald",]
Donald
```

```
## # A tibble: 52 x 5
##     year sex   name       n       prop
##    <dbl> <chr> <chr>  <int>      <dbl>
##  1  1938 F     Donald    79 0.0000692
##  2  1936 M     Donald 28635 0.0269
##  3  2016 M     Donald   624 0.000309
##  4  2004 F     Donald     8 0.00000397
##  5  1942 F     Donald    76 0.0000547
##  6  1907 M     Donald   616 0.00388
##  7  1902 M     Donald   366 0.00276
##  8  1971 F     Donald    67 0.0000382
##  9  1951 M     Donald 27918 0.0146
## 10  1965 M     Donald 16040 0.00846
## # ... with 42 more rows
```

```
c1<-sum(Donald$n)
c1
```

```
## [1] 331949
```

```
Hilary <- sample[sample$name=="Hilary",]
Hilary
```

```
## # A tibble: 61 x 5
##     year sex   name       n       prop
##    <dbl> <chr> <chr>  <int>      <dbl>
##  1  1942 F     Hilary    36 0.0000259
##  2  1915 M     Hilary    36 0.0000409
##  3  2010 F     Hilary    80 0.0000408
##  4  1936 F     Hilary     9 0.00000835
##  5  1947 F     Hilary    62 0.0000341
##  6  1984 M     Hilary    10 0.00000533
##  7  1951 F     Hilary   103 0.0000558
##  8  1883 M     Hilary     6 0.0000533
##  9  1917 M     Hilary    44 0.0000459
## 10  1987 M     Hilary     8 0.0000041
## # ... with 51 more rows
```

```
c2<-sum(Hilary$n)
```

```
Hillary <- sample[sample$name=="Hillary",]
Hillary
```

```
## # A tibble: 44 x 5
##     year sex   name        n       prop
##    <dbl> <chr> <chr>   <int>      <dbl>
##  1  1962 M     Hillary    17 0.00000809
##  2  1977 F     Hillary   442 0.000269
##  3  1994 M     Hillary     7 0.00000343
##  4  2012 F     Hillary   157 0.0000811
##  5  1950 M     Hillary    20 0.0000110
##  6  1974 F     Hillary   313 0.000200
##  7  1989 M     Hillary    14 0.00000668
##  8  1915 M     Hillary    18 0.0000204
##  9  2015 F     Hillary   137 0.0000704
## 10  1970 F     Hillary   258 0.000141
## # ... with 34 more rows
```

```
c3<-sum(Hillary$n)
c3
```

```
## [1] 9276
```

```
Joe <- sample[sample$name=="Joe",]
Joe
```

```
## # A tibble: 55 x 5
##     year sex   name      n      prop
##    <dbl> <chr> <chr> <int>     <dbl>
##  1  1952 F     Joe     174 0.0000915
##  2  1906 F     Joe      42 0.000134
##  3  1896 F     Joe      12 0.0000476
##  4  1902 F     Joe      30 0.000107
##  5  1914 M     Joe    3712 0.00543
##  6  1957 F     Joe     172 0.000082
##  7  2006 M     Joe     808 0.000369
##  8  1995 M     Joe    1077 0.000536
##  9  1880 M     Joe     731 0.00617
## 10  1974 F     Joe      48 0.0000306
## # ... with 45 more rows
```

```
c4<-sum(Joe$n)
c4
```

```
## [1] 60958
```

```
Barrack <- sample[sample$name=="Barrack",]
Barrack
```

```
## # A tibble: 0 x 5
## # ... with 5 variables: year <dbl>, sex <chr>, name <chr>, n <int>,
## #   prop <dbl>
```

```
c5<-sum(Barrack$n)
c5
```

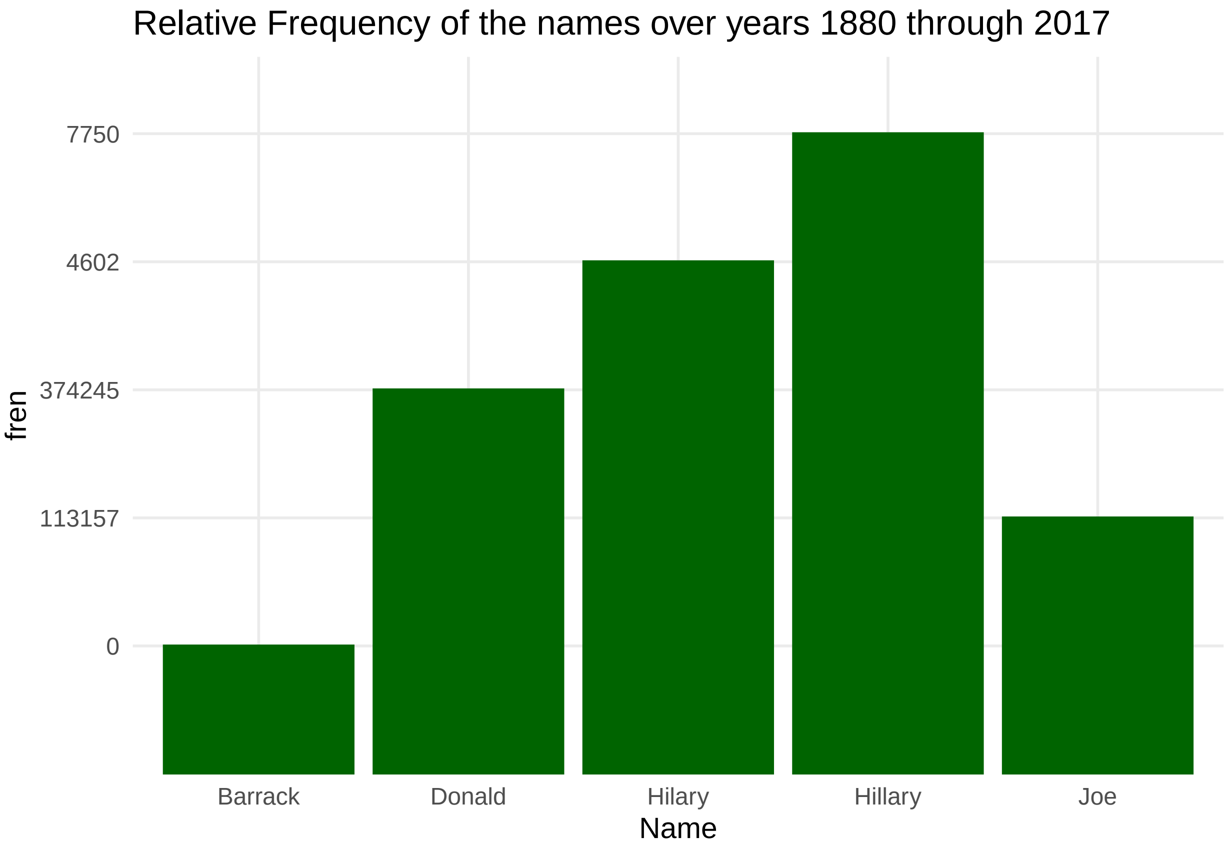
```
## [1] 0
```

```
fren<-c(374245,4602,7750,113157,0)
fren
```

```
## [1] 374245   4602   7750 113157      0
```

```
Name<-c("Donald","Hilary","Hillary","Joe","Barrack")
frenquen<-data.frame(cbind(Name,fren))
```

```
ggplot(frenquen,aes(y=fren,x=Name))+
  geom_bar(stat="identity",color="darkgreen",fill="darkgreen")+theme_minimal()+
  ggtitle("Relative Frequency of the names over years 1880 through 2017")
```


Relative Frequency of the names over years 1880 through 2017
fren
Barrack
Donald
Hilary
Hillary
Joe
Name