Tidyverse Exercise

Kerui Cao

10/1/2019

Problem 1

How many continents are included in the data set?

The numbers of countries is 5.

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

There are 142 contries included. So 28.4 contries per continents. More specific and accurate infomation is listed below"

Table 1: Number of countries included for each cintinents

continent	Num_of_contries
Oceania	2
Americas	25
Europe	30
Asia	33
Africa	52

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.

Table 2: Total Population and GDP per Capita

Year	Africa		American		Asia	
Year	Total Population	GDP Per Capita	Total Population.1	GDP Per Capita.1	Total Population.2	GDP Per Capita.2
1952	237640501	1311.221	345152446	8528.043	1395357351	806.3599
1957	264837738	1444.952	386953916	9097.793	1562780599	998.1089
1962	296516865	1540.599	433270254	9760.252	1696357182	1169.8696
1967	335289489	1774.847	480746623	11329.644	1905662900	1465.8422
1972	379879541	2063.171	529384210	12663.731	2150972248	1908.3136
1977	433061021	2244.799	578067699	14015.892	2384513556	2211.5561
1982	499348587	2295.192	630290920	14410.568	2610135582	2458.1706
1987	574834110	2180.764	682753971	16091.001	2871220762	2778.9215
1992	659081517	2071.615	739274104	16566.921	3133292191	3234.3988
1997	743832984	2098.865	796900410	18279.991	3383285500	3780.5626
2002	833723916	2201.907	849772762	19453.690	3601802203	4196.6205
2007	929539692	2560.930	898871184	21602.746	3811953827	5432.3717

Table 3: Total Population and GDP per Capita

Year	Europe		Oceania	
Year	Total Population	GDP Per Capita	Total Population.1	GDP Per Capita.1
1952	418120846	6096.659	10686006	10136.10
1957	437890351	7535.414	11941976	11191.92
1962	460355155	9057.226	13283518	12396.78
1967	481178958	10808.867	14600414	14514.50
1972	500635059	13104.843	16106100	16653.58
1977	517164531	14813.517	17239000	17948.57
1982	531266901	15782.128	18394850	19155.05
1987	543094160	17483.569	19574415	21400.54
1992	558142797	18420.192	20919651	22593.03
1997	568944148	20227.307	22241430	26014.90
2002	578223869	22644.282	23454829	29438.45
2007	586098529	25244.048	24549947	32884.56

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

Table 4: Summary for GDP per Capita in 1952

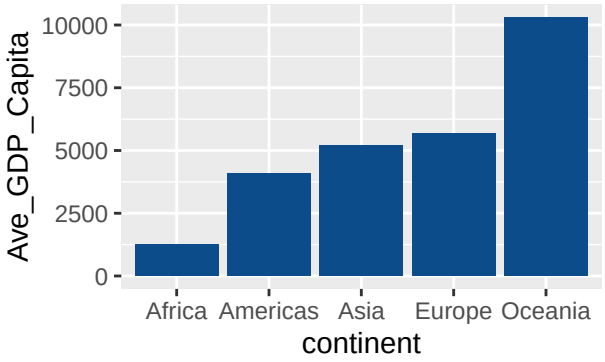
continent	Ave_GDP_Capita	Max_GDP_Capita	Min_GDP_Capita	var_GDP_Capita
Africa	1252.572	4725.296	298.8462	966194.9
Americas	4079.063	13990.482	1397.7171	9010368.1
Asia	5195.484	108382.353	331.0000	347259157.6
Europe	5661.057	14734.233	973.5332	9697372.8
Oceania	10298.086	10556.576	10039.5956	133634.2

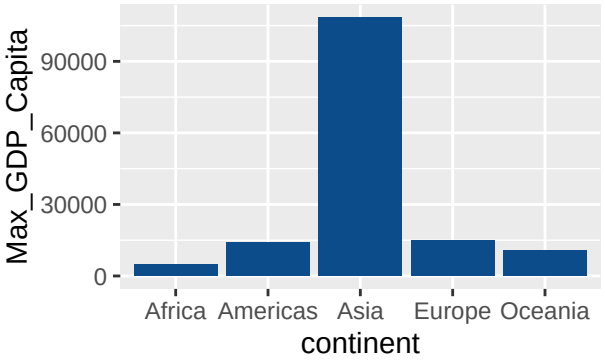
Table 5: Summary for GDP per Capita in 2007

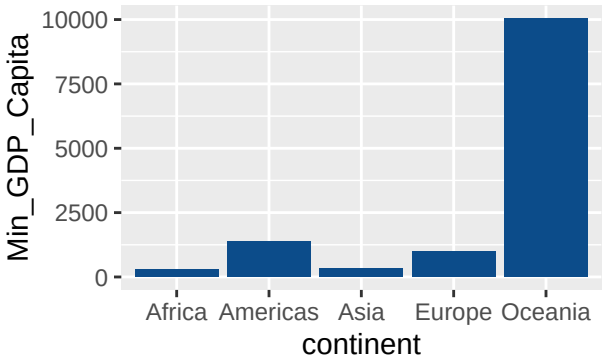
continent	Ave_GDP_Capita	Max_GDP_Capita	Min_GDP_Capita	var_GDP_Capita
Africa	3089.033	13206.48	277.5519	13091107
Americas	11003.032	42951.65	1201.6372	94346435
Asia	12473.027	47306.99	944.0000	200362251
Europe	25054.482	49357.19	5937.0295	139248020
Oceania	29810.188	34435.37	25185.0091	42784565

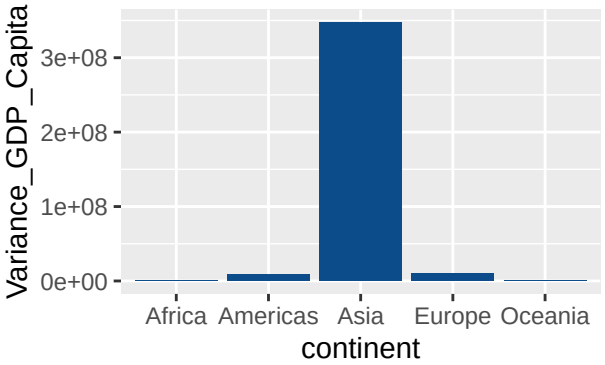
Product a plot that summarizes the same data as the table. There should be two plots per continent.

Summary for year 1952

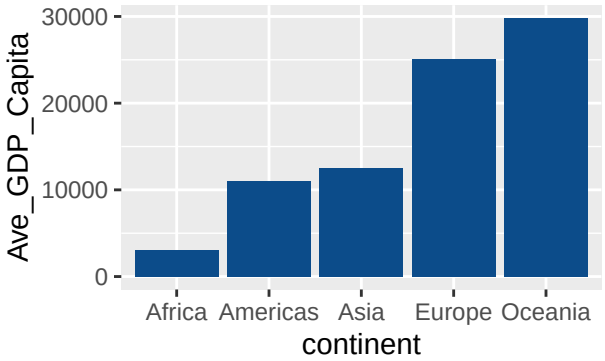


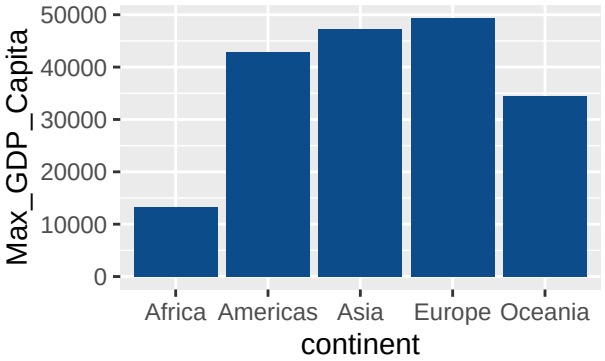


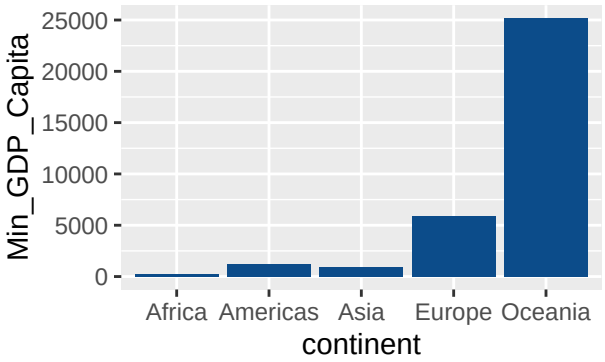


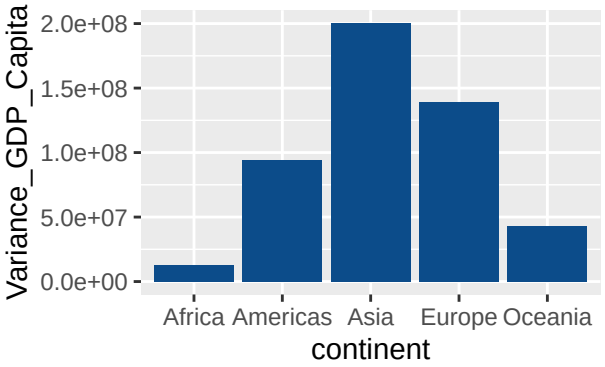


Summary for year 2007









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

Table 6: Countries had periods of negative population growth

Country	# of year of negative pop growth	Country.1	# of year of negative pop growth.1	Country.2	# of year of negative pop growth.2
Afghanistan	1	Montenegro	1	Germany	2
Cambodia	1	Portugal	1	Ireland	2
Croatia	1	Rwanda	1	Poland	2
Equatorial Guinea	1	Serbia	1	Slovenia	2
Guinea-Bissau	1	Somalia	1	Czech Republic	3
Kuwait	1	South Africa	1	Romania	3
Lebanon	1	Switzerland	1	Bulgaria	4
Lesotho	1	West Bank and Gaza	1	Trinidad and Tobago	4
Liberia	1	Bosnia and Herzegovina	2	Hungary	5

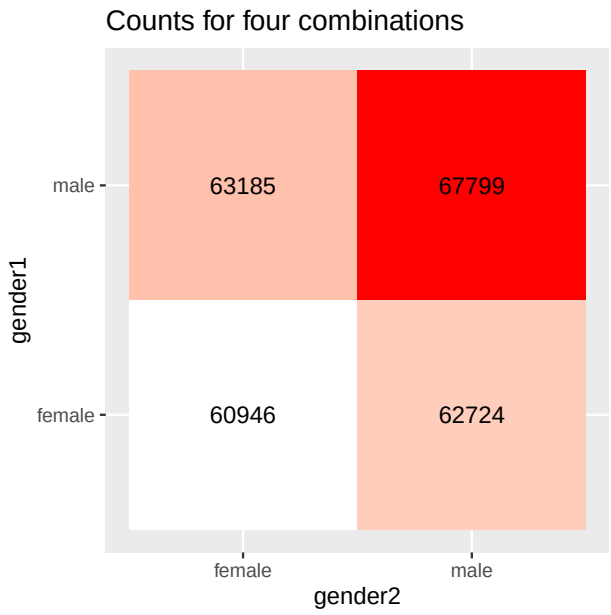
Which countries in the dataset have had the highest rate of growth in per capita GDP? Illustrate your answer with a table or plot.

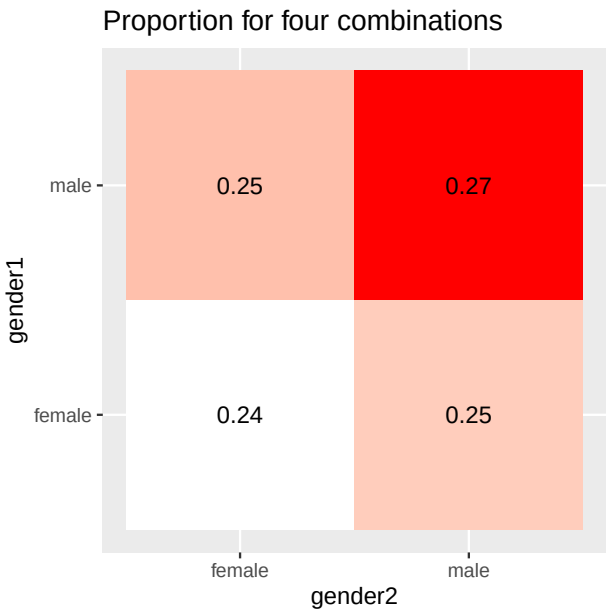
Table 7: Log Growth Rate

country	Max_GR
Libya	1.0218229
Equatorial Guinea	1.0068965
Oman	0.8105458
Cambodia	0.6482633
Gabon	0.6456260
Bosnia and Herzegovina	0.6267517
Botswana	0.6224566
Angola	0.5480056
Singapore	0.5465899
Korea, Dem. Rep.	0.5463120

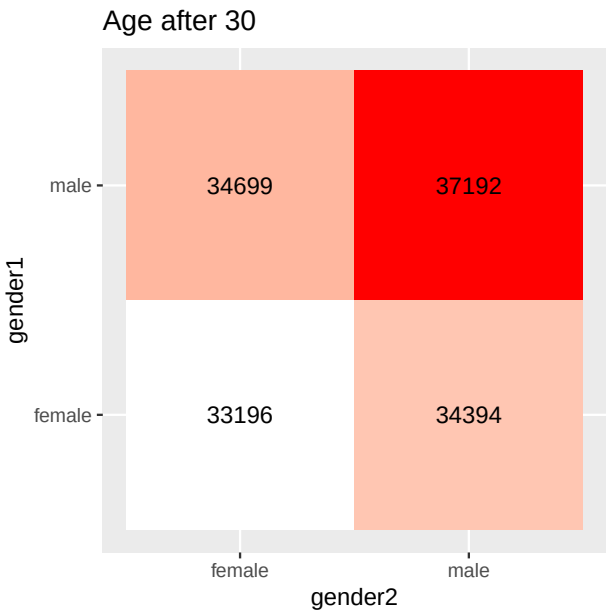
Problem2

Frequency of four combinations.



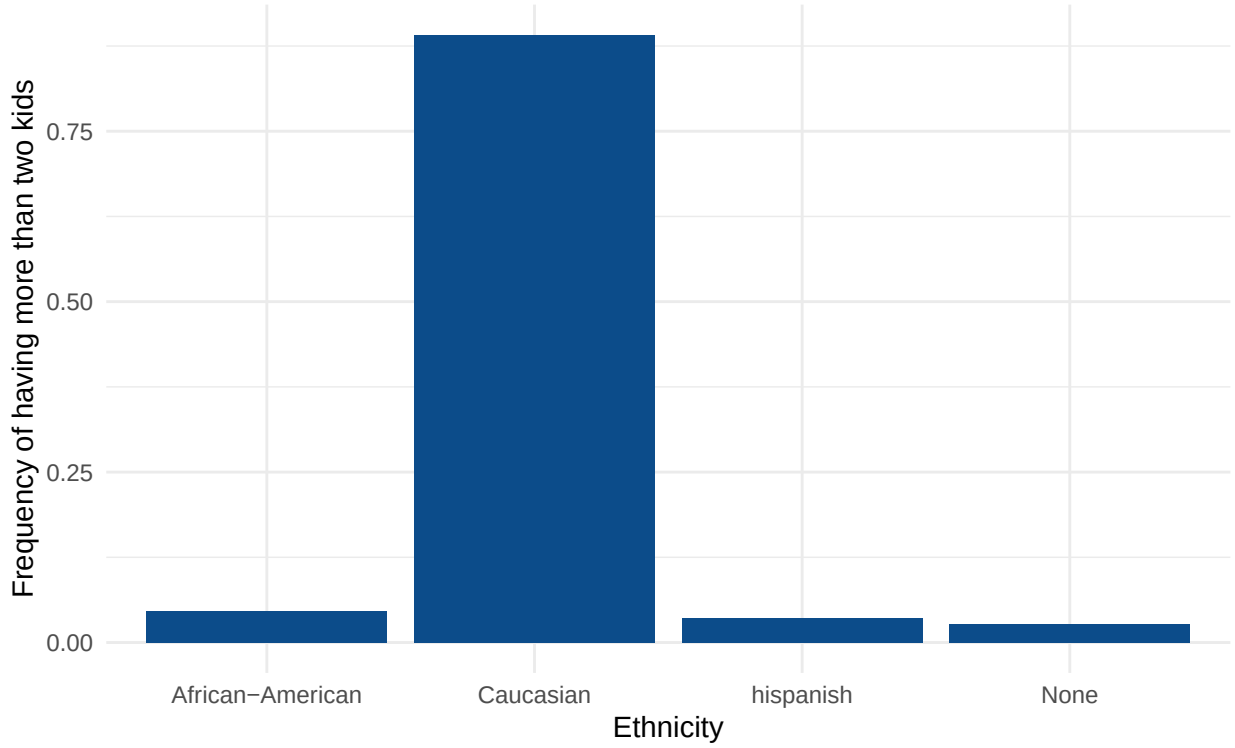


Are the frequencies different for women in their 20s and wemen who are older than 29?



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

According to the description of Fertility, we can know that there are conflicts in the ethnicity, so we just choose those records that don‘t conflict.



Problem 3

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

```
q3 = sum(str_count(rownames(mtcars),"e"))
```

There are 25 cars whose names contain "e".

How many cars in mtcars have the brand Merc?

```
q4 = sum(str_detect(rownames(mtcars),"Merc"))
```

There are 7 cars in mtcars have the brand Merc.

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

4 cars in mpg have the brand Merc.

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

Problem 4

Sample 500000 records from "babynames".

```
R_in = sample(1:1924665,500000,replace = F)
sam_da = babynames[R_in,]
```

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

Table 8: Most popular name each year

year	sex	name	n	prop
1880	F	Emma	2003	0.02052149
1880	F	Minnie	1746	0.01788843
1880	F	Margaret	1578	0.01616720
1880	F	Alice	1414	0.01448696
1880	F	Sarah	1288	0.01319605
1880	M	Charles	5348	0.04516892
1880	M	Clarence	730	0.00616554
1880	M	Richard	728	0.00614865
1880	M	Peter	496	0.00418919
1880	M	Willie	476	0.00402027
1920	F	Helen	35097	0.02821214
1920	F	Mildred	18058	0.01451562
1920	F	Betty	14017	0.01126733
1920	F	Marie	12743	0.01024325
1920	F	Florence	10732	0.00862674
1920	M	John	56913	0.05170069
1920	M	Robert	48678	0.04421988
1920	M	Harold	13679	0.01242623
1920	M	Clarence	7222	0.00656058
1920	M	Warren	5492	0.00498902
1960	F	Mary	51474	0.02474901
1960	F	Susan	39200	0.01884759
1960	F	Linda	37314	0.01794079
1960	F	Karen	36376	0.01748980
1960	F	Donna	34133	0.01641135
1960	M	Michael	84183	0.03887111
1960	M	Timothy	30484	0.01407584
1960	M	Jeffrey	28831	0.01331258
1960	M	Brian	21994	0.01015563
1960	M	Ronald	21700	0.01001987
2000	F	Emily	25953	0.01300980
2000	F	Ashley	17997	0.00902159
2000	F	Alexis	17629	0.00883712
2000	F	Samantha	17266	0.00865515
2000	F	Lauren	14175	0.00710569
2000	M	Joseph	22825	0.01093470
2000	M	Alexander	17282	0.00827923
2000	M	Jonathan	16882	0.00808761
2000	M	Austin	15944	0.00763824
2000	M	Samuel	14170	0.00678838

What names overlap boys and girls?

```
hh = sam_da %>% group_by(year,name) %>% summarise(cou = length(sex)) %>% arrange(desc(cou)) %>%
  filter(cou>1)
unique(hh$name)[1:10]
```

```
##  [1] "Charles" "Connie"  "Dee"     "Dora"    "Fay"     "Johnnie" "Minnie"
##  [8] "Vivian"  "Allie"   "Essie"
```

There are over 4000 names that overlap boys and girls, so here I displayed 10 of them.

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

```
l1 = sam_da %>% filter(year>1999)
l1 = unique(l1$name)
l2 = sam_da %>% filter(year<1900)
l2 = unique(l2$name)
Int = intersect(l1,l2)
Int[1:10]
```

```
##  [1] "Hurley" "Ivy"    "Zellie" "Allene" "Ruthie" "Addie"  "Lula"
##  [8] "Colin"  "Maria"  "Lelah"
```

There are 2562 names that were used in the 19th century but have not been used in the 21th centure, so here I listed 10 of them.

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

name	count	frequency
Donald	9133	0.36
Hilary	1403	0.06
Hillary	1837	0.07
Joe	13131	0.51