Homework 2

Bhoomika Khatri

2024-02-26

Data Cleaning in R

You will use R to clean some data.

1. Load the library dplyr

```
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

2. Read in the .csv file gapminder dataset

It is in the subfolder of this folder.

```
getwd()
```

```
[1] "/mnt/pan/courses/csds133/bxk421/24sp-csds133-bxk421/1-assignments/LabExcercise-2"
```

```
setwd("./datafiles")
gapMinderData <- read.csv("gapminder.csv")
```

3. Do some preliminary data exploration with str and summary functions.

Comment your code with 2 interesting things you see

```
str(gapMinderData) #there are 5 years in between the years of data for each country
```

```
'data.frame':   1704 obs. of  7 variables:
 $ X        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ country  : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ year     : int  1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
 $ pop      : num  8425333 9240934 10267083 11537966 13079460 ...
 $ continent: chr  "Asia" "Asia" "Asia" "Asia" ...
 $ lifeExp  : num  28.8 30.3 32 34 36.1 ...
 $ gdpPercap: num  779 821 853 836 740 ...
```

```
the_summary <- summary(gapMinderData) #there are about 12 instances of each country if it
```

4. Clean the data by converting the character class columns to factors

```
gapMinderData$country <- as.factor(gapMinderData$country)
gapMinderData$continent <- as.factor(gapMinderData$continent)
```

5. Use the functions in dplyr such as filter(), select(), mutate(), etc to explore your data and identify something interesting in your the gapminder data frame. Describe what you identified. (2)

```
gapMinderData %>% dplyr::select(continent, country, lifeExp, year) %>% filter(lifeExp <= 3
```

```
   continent      country lifeExp year
1       Asia Afghanistan  28.801 1952
2     Africa       Gambia  30.000 1952
3     Africa       Rwanda  23.599 1992
```

- There are only 3 countries that have a life expectancy of under (or equal) to 30 in the span of 1952 - 2007, and all of them are from 1952. This means that between 1952 and 197, all countries had a life expectancy of over 30 increased their life expectancy.

6. Use the functions in dplyr such as filter(), select(), mutate(), etc to explore your data and identify something interesting in your the gapminder data frame. Describe what you identified. (2)

```
gapMinderData %>% arrange(desc(gdpPercap)) %>% group_by(continent) %>% summarize(Ave = ave
```

```
Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
dplyr 1.1.0.
i Please use `reframe()` instead.
i When switching from `summarise()` to `reframe()`, remember that `reframe()`
  always returns an ungrouped data frame and adjust accordingly.
```

```
`summarise()` has grouped output by 'continent'. You can override using the
`.groups` argument.
```

```
# A tibble: 5 x 3
# Groups:   continent [5]
  continent    Ave     Max
  <fct>      <dbl>   <dbl>
1 Africa     2194.  21951.
2 Oceania   18622.  34435.
3 Americas   7136.  42952.
4 Europe    14469.  49357.
5 Asia       7902. 113523.
```

- Surprisingly, Oceania had the highest average GDP and Asia had the highest GDP. I would've expected that Europe or the Americans would have the highest GDP due to the western countries, but it didn't end up that way.

3

7. Use the functions in dplyr such as filter(), select(), mutate(), etc to explore your data and identify something interesting in your the gapminder data frame. Describe what you identified. (2)

```
gapMinderData  %>% arrange(desc(pop),year) %>% group_by(country) %>% summarize(change = ma
```

```
# A tibble: 142 x 2
   country               change
   <fct>                  <dbl>
 1 Sao Tome and Principe 139568
 2 Iceland               153969
 3 Montenegro            306396
 4 Equatorial Guinea     358526
 5 Djibouti              433225
 6 Slovenia              522094
 7 Trinidad and Tobago   528486
 8 Reunion               540394
 9 Comoros               557024
10 Bahrain               588126
# i 132 more rows
```

- This data set shows that most countries had a drastic change from the minimum population size of the country to the maximum size of the data set. The country with the largest change is China and the country with the smallest change is Sao Tome and Principe. Coincidentally, these are also the least and largest populations (respectively).

8. Use the functions in dplyr such as filter(), select(), mutate(), etc to explore your data and identify something interesting in your the gapminder data frame. Describe what you identified. (2)

```
gapMinderData %>% dplyr::select(continent, country, lifeExp, year) %>% filter(lifeExp >= 7
```

```
   continent        country lifeExp year
1    Europe         Denmark  70.780 1952
2    Europe         Iceland  72.490 1952
3    Europe     Netherlands  72.130 1952
4    Europe          Norway  72.670 1952
5    Europe          Sweden  71.860 1952
6   Oceania       Australia  70.330 1957
```

4

7	Europe	Denmark	71.810	1957
8	Europe	Iceland	73.470	1957
9	Europe	Netherlands	72.990	1957
10	Oceania	New Zealand	70.260	1957
11	Europe	Norway	73.440	1957
12	Europe	Sweden	72.490	1957
13	Europe	Switzerland	70.560	1957
14	Europe	United Kingdom	70.420	1957
15	Oceania	Australia	70.930	1962
16	Europe	Belgium	70.250	1962
17	Americas	Canada	71.300	1962
18	Europe	Denmark	72.350	1962
19	Europe	France	70.510	1962
20	Europe	Germany	70.300	1962
21	Europe	Iceland	73.680	1962
22	Europe	Ireland	70.290	1962
23	Europe	Netherlands	73.230	1962
24	Oceania	New Zealand	71.240	1962
25	Europe	Norway	73.470	1962
26	Europe	Slovak Republic	70.330	1962
27	Europe	Sweden	73.370	1962
28	Europe	Switzerland	71.320	1962
29	Europe	United Kingdom	70.760	1962
30	Americas	United States	70.210	1962
31	Oceania	Australia	71.100	1967
32	Europe	Austria	70.140	1967
33	Europe	Belgium	70.940	1967
34	Europe	Bulgaria	70.420	1967
35	Americas	Canada	72.130	1967
36	Europe	Czech Republic	70.380	1967
37	Europe	Denmark	72.960	1967
38	Europe	France	71.550	1967
39	Europe	Germany	70.800	1967
40	Europe	Greece	71.000	1967
41	Asia	Hong Kong, China	70.000	1967
42	Europe	Iceland	73.730	1967
43	Europe	Ireland	71.080	1967
44	Asia	Israel	70.750	1967
45	Europe	Italy	71.060	1967
46	Asia	Japan	71.430	1967
47	Europe	Netherlands	73.820	1967
48	Oceania	New Zealand	71.520	1967
49	Europe	Norway	74.080	1967

```
50   Americas        Puerto Rico  71.100 1967
51     Europe     Slovak Republic  70.980 1967
52     Europe               Spain  71.440 1967
53     Europe              Sweden  74.160 1967
54     Europe         Switzerland  72.770 1967
55     Europe      United Kingdom  71.360 1967
56   Americas       United States  70.760 1967
57    Oceania           Australia  71.930 1972
58     Europe             Austria  70.630 1972
59     Europe             Belgium  71.440 1972
60     Europe            Bulgaria  70.900 1972
61   Americas              Canada  72.880 1972
62   Americas                Cuba  70.723 1972
63     Europe      Czech Republic  70.290 1972
64     Europe             Denmark  73.470 1972
65     Europe             Finland  70.870 1972
66     Europe              France  72.380 1972
67     Europe             Germany  71.000 1972
68     Europe              Greece  72.340 1972
69       Asia    Hong Kong, China  72.000 1972
70     Europe             Iceland  74.460 1972
71     Europe             Ireland  71.280 1972
72       Asia              Israel  71.630 1972
73     Europe               Italy  72.190 1972
74       Asia               Japan  73.420 1972
75     Europe          Montenegro  70.636 1972
76     Europe         Netherlands  73.750 1972
77    Oceania         New Zealand  71.890 1972
78     Europe              Norway  74.340 1972
79     Europe              Poland  70.850 1972
80   Americas         Puerto Rico  72.160 1972
81     Europe     Slovak Republic  70.350 1972
82     Europe               Spain  73.060 1972
83     Europe              Sweden  74.720 1972
84     Europe         Switzerland  73.780 1972
85     Europe      United Kingdom  72.010 1972
86   Americas       United States  71.340 1972
87    Oceania           Australia  73.490 1977
88     Europe             Austria  72.170 1977
89     Europe             Belgium  72.800 1977
90     Europe            Bulgaria  70.810 1977
91   Americas              Canada  74.210 1977
92   Americas          Costa Rica  70.750 1977
```

```
93    Europe                 Croatia  70.640 1977
94  Americas                    Cuba  72.649 1977
95    Europe          Czech Republic  70.710 1977
96    Europe                 Denmark  74.690 1977
97    Europe                 Finland  72.520 1977
98    Europe                  France  73.830 1977
99    Europe                 Germany  72.500 1977
100   Europe                  Greece  73.680 1977
101     Asia        Hong Kong, China  73.600 1977
102   Europe                 Iceland  76.110 1977
103   Europe                 Ireland  72.030 1977
104     Asia                  Israel  73.060 1977
105   Europe                   Italy  73.480 1977
106 Americas                 Jamaica  70.110 1977
107     Asia                   Japan  75.380 1977
108   Europe              Montenegro  73.066 1977
109   Europe             Netherlands  75.240 1977
110  Oceania             New Zealand  72.220 1977
111   Europe                  Norway  75.370 1977
112   Europe                  Poland  70.670 1977
113   Europe                Portugal  70.410 1977
114 Americas             Puerto Rico  73.440 1977
115   Europe                  Serbia  70.300 1977
116     Asia               Singapore  70.795 1977
117   Europe         Slovak Republic  70.450 1977
118   Europe                Slovenia  70.970 1977
119   Europe                   Spain  74.390 1977
120   Europe                  Sweden  75.440 1977
121   Europe             Switzerland  75.390 1977
122     Asia                  Taiwan  70.590 1977
123   Europe          United Kingdom  72.760 1977
124 Americas           United States  73.380 1977
125   Europe                 Albania  70.420 1982
126  Oceania               Australia  74.740 1982
127   Europe                 Austria  73.180 1982
128   Europe                 Belgium  73.930 1982
129   Europe  Bosnia and Herzegovina  70.690 1982
130   Europe                Bulgaria  71.080 1982
131 Americas                  Canada  75.760 1982
132 Americas                   Chile  70.565 1982
133 Americas              Costa Rica  73.450 1982
134   Europe                 Croatia  70.460 1982
135 Americas                    Cuba  73.717 1982
```

```
136     Europe                Czech Republic  70.960 1982
137     Europe                      Denmark   74.630 1982
138     Europe                      Finland   74.550 1982
139     Europe                       France   74.890 1982
140     Europe                      Germany   73.800 1982
141     Europe                       Greece   75.240 1982
142       Asia            Hong Kong, China    75.450 1982
143     Europe                      Iceland   76.990 1982
144     Europe                      Ireland   73.100 1982
145       Asia                       Israel   74.450 1982
146     Europe                        Italy   74.980 1982
147   Americas                      Jamaica   71.210 1982
148       Asia                        Japan   77.110 1982
149       Asia                       Kuwait   71.309 1982
150     Europe                   Montenegro   74.101 1982
151     Europe                  Netherlands   76.050 1982
152    Oceania                  New Zealand   73.840 1982
153     Europe                       Norway   75.970 1982
154   Americas                       Panama   70.472 1982
155     Europe                       Poland   71.320 1982
156     Europe                     Portugal   72.770 1982
157   Americas                  Puerto Rico   73.750 1982
158     Europe                       Serbia   70.162 1982
159       Asia                    Singapore   71.760 1982
160     Europe              Slovak Republic   70.800 1982
161     Europe                     Slovenia   71.063 1982
162     Europe                        Spain   76.300 1982
163     Europe                       Sweden   76.420 1982
164     Europe                  Switzerland   76.210 1982
165       Asia                       Taiwan   72.160 1982
166     Europe               United Kingdom   74.040 1982
167   Americas                United States   74.650 1982
168   Americas                      Uruguay   70.805 1982
169     Europe                      Albania   72.000 1987
170   Americas                    Argentina   70.774 1987
171    Oceania                    Australia   76.320 1987
172     Europe                      Austria   74.940 1987
173       Asia                      Bahrain   70.750 1987
174     Europe                      Belgium   75.350 1987
175     Europe  Bosnia and Herzegovina       71.140 1987
176     Europe                     Bulgaria   71.340 1987
177   Americas                       Canada   76.860 1987
178   Americas                        Chile   72.492 1987
```

```
179   Americas              Costa Rica  74.752 1987
180     Europe                 Croatia  71.520 1987
181   Americas                    Cuba  74.174 1987
182     Europe          Czech Republic  71.580 1987
183     Europe                 Denmark  74.800 1987
184     Europe                 Finland  74.830 1987
185     Europe                  France  76.340 1987
186     Europe                 Germany  74.847 1987
187     Europe                  Greece  76.670 1987
188       Asia        Hong Kong, China  76.200 1987
189     Europe                 Iceland  77.230 1987
190     Europe                 Ireland  74.360 1987
191       Asia                  Israel  75.600 1987
192     Europe                   Italy  76.420 1987
193   Americas                 Jamaica  71.770 1987
194       Asia                   Japan  78.670 1987
195       Asia        Korea, Dem. Rep.  70.647 1987
196       Asia                  Kuwait  74.174 1987
197     Europe              Montenegro  74.865 1987
198     Europe             Netherlands  76.830 1987
199    Oceania             New Zealand  74.320 1987
200     Europe                  Norway  75.890 1987
201   Americas                  Panama  71.523 1987
202     Europe                  Poland  70.980 1987
203     Europe                Portugal  74.060 1987
204   Americas             Puerto Rico  74.630 1987
205     Africa                 Reunion  71.913 1987
206     Europe                  Serbia  71.218 1987
207       Asia               Singapore  73.560 1987
208     Europe         Slovak Republic  71.080 1987
209     Europe                Slovenia  72.250 1987
210     Europe                   Spain  76.900 1987
211     Europe                  Sweden  77.190 1987
212     Europe             Switzerland  77.410 1987
213       Asia                  Taiwan  73.400 1987
214     Europe          United Kingdom  75.007 1987
215   Americas           United States  75.020 1987
216   Americas                 Uruguay  71.918 1987
217   Americas               Venezuela  70.190 1987
218     Europe                 Albania  71.581 1992
219   Americas               Argentina  71.868 1992
220    Oceania               Australia  77.560 1992
221     Europe                 Austria  76.040 1992
```

```
222       Asia                  Bahrain  72.601 1992
223     Europe                  Belgium  76.460 1992
224     Europe Bosnia and Herzegovina  72.178 1992
225     Europe                 Bulgaria  71.190 1992
226   Americas                   Canada  77.950 1992
227   Americas                    Chile  74.126 1992
228   Americas               Costa Rica  75.713 1992
229     Europe                  Croatia  72.527 1992
230   Americas                     Cuba  74.414 1992
231     Europe           Czech Republic  72.400 1992
232     Europe                  Denmark  75.330 1992
233     Europe                  Finland  75.700 1992
234     Europe                   France  77.460 1992
235     Europe                  Germany  76.070 1992
236     Europe                   Greece  77.030 1992
237       Asia         Hong Kong, China  77.601 1992
238     Europe                  Iceland  78.770 1992
239     Europe                  Ireland  75.467 1992
240       Asia                   Israel  76.930 1992
241     Europe                    Italy  77.440 1992
242   Americas                  Jamaica  71.766 1992
243       Asia                    Japan  79.360 1992
244       Asia              Korea, Rep.  72.244 1992
245       Asia                   Kuwait  75.190 1992
246       Asia                 Malaysia  70.693 1992
247   Americas                   Mexico  71.455 1992
248     Europe               Montenegro  75.435 1992
249     Europe              Netherlands  77.420 1992
250    Oceania              New Zealand  76.330 1992
251     Europe                   Norway  77.320 1992
252       Asia                     Oman  71.197 1992
253   Americas                   Panama  72.462 1992
254     Europe                   Poland  70.990 1992
255     Europe                 Portugal  74.860 1992
256   Americas              Puerto Rico  73.911 1992
257     Africa                  Reunion  73.615 1992
258     Europe                   Serbia  71.659 1992
259       Asia                Singapore  75.788 1992
260     Europe          Slovak Republic  71.380 1992
261     Europe                 Slovenia  73.640 1992
262     Europe                    Spain  77.570 1992
263       Asia                Sri Lanka  70.379 1992
264     Europe                   Sweden  78.160 1992
```

```
265     Europe                Switzerland  78.030 1992
266       Asia                     Taiwan  74.260 1992
267     Africa                    Tunisia  70.001 1992
268     Europe             United Kingdom  76.420 1992
269   Americas              United States  76.090 1992
270   Americas                    Uruguay  72.752 1992
271   Americas                  Venezuela  71.150 1992
272     Europe                    Albania  72.950 1997
273   Americas                  Argentina  73.275 1997
274    Oceania                  Australia  78.830 1997
275     Europe                    Austria  77.510 1997
276       Asia                    Bahrain  73.925 1997
277     Europe                    Belgium  77.530 1997
278     Europe Bosnia and Herzegovina  73.244 1997
279     Europe                   Bulgaria  70.320 1997
280   Americas                     Canada  78.610 1997
281   Americas                      Chile  75.816 1997
282       Asia                      China  70.426 1997
283   Americas                   Colombia  70.313 1997
284   Americas                 Costa Rica  77.260 1997
285     Europe                    Croatia  73.680 1997
286   Americas                       Cuba  76.151 1997
287     Europe             Czech Republic  74.010 1997
288     Europe                    Denmark  76.110 1997
289   Americas                    Ecuador  72.312 1997
290     Europe                    Finland  77.130 1997
291     Europe                     France  78.640 1997
292     Europe                    Germany  77.340 1997
293     Europe                     Greece  77.869 1997
294       Asia           Hong Kong, China  80.000 1997
295     Europe                    Hungary  71.040 1997
296     Europe                    Iceland  78.950 1997
297     Europe                    Ireland  76.122 1997
298       Asia                     Israel  78.269 1997
299     Europe                      Italy  78.820 1997
300   Americas                    Jamaica  72.262 1997
301       Asia                      Japan  80.690 1997
302       Asia                Korea, Rep.  74.647 1997
303       Asia                     Kuwait  76.156 1997
304       Asia                    Lebanon  70.265 1997
305     Africa                      Libya  71.555 1997
306       Asia                   Malaysia  71.938 1997
307     Africa                  Mauritius  70.736 1997
```

```
308    Americas                      Mexico  73.670 1997
309     Europe                  Montenegro  75.445 1997
310     Europe                 Netherlands  78.030 1997
311    Oceania                 New Zealand  77.550 1997
312     Europe                      Norway  78.320 1997
313       Asia                        Oman  72.499 1997
314    Americas                      Panama  73.738 1997
315     Europe                      Poland  72.750 1997
316     Europe                    Portugal  75.970 1997
317    Americas                 Puerto Rico  74.917 1997
318     Africa                     Reunion  74.772 1997
319       Asia                Saudi Arabia  70.533 1997
320     Europe                      Serbia  72.232 1997
321       Asia                   Singapore  77.158 1997
322     Europe              Slovak Republic  72.710 1997
323     Europe                    Slovenia  75.130 1997
324     Europe                       Spain  78.770 1997
325       Asia                   Sri Lanka  70.457 1997
326     Europe                      Sweden  79.390 1997
327     Europe                 Switzerland  79.370 1997
328       Asia                       Syria  71.527 1997
329       Asia                      Taiwan  75.250 1997
330     Africa                     Tunisia  71.973 1997
331     Europe              United Kingdom  77.218 1997
332    Americas              United States  76.810 1997
333    Americas                     Uruguay  74.223 1997
334    Americas                   Venezuela  72.146 1997
335       Asia                     Vietnam  70.672 1997
336       Asia          West Bank and Gaza  71.096 1997
337     Europe                     Albania  75.651 2002
338     Africa                     Algeria  70.994 2002
339    Americas                   Argentina  74.340 2002
340    Oceania                   Australia  80.370 2002
341     Europe                     Austria  78.980 2002
342       Asia                     Bahrain  74.795 2002
343     Europe                     Belgium  78.320 2002
344     Europe  Bosnia and Herzegovina  74.090 2002
345    Americas                      Brazil  71.006 2002
346     Europe                    Bulgaria  72.140 2002
347    Americas                      Canada  79.770 2002
348    Americas                       Chile  77.860 2002
349       Asia                       China  72.028 2002
350    Americas                    Colombia  71.682 2002
```

351	Americas	Costa Rica	78.123	2002
352	Europe	Croatia	74.876	2002
353	Americas	Cuba	77.158	2002
354	Europe	Czech Republic	75.510	2002
355	Europe	Denmark	77.180	2002
356	Americas	Dominican Republic	70.847	2002
357	Americas	Ecuador	74.173	2002
358	Americas	El Salvador	70.734	2002
359	Europe	Finland	78.370	2002
360	Europe	France	79.590	2002
361	Europe	Germany	78.670	2002
362	Europe	Greece	78.256	2002
363	Asia	Hong Kong, China	81.495	2002
364	Europe	Hungary	72.590	2002
365	Europe	Iceland	80.500	2002
366	Europe	Ireland	77.783	2002
367	Asia	Israel	79.696	2002
368	Europe	Italy	80.240	2002
369	Americas	Jamaica	72.047	2002
370	Asia	Japan	82.000	2002
371	Asia	Jordan	71.263	2002
372	Asia	Korea, Rep.	77.045	2002
373	Asia	Kuwait	76.904	2002
374	Asia	Lebanon	71.028	2002
375	Africa	Libya	72.737	2002
376	Asia	Malaysia	73.044	2002
377	Africa	Mauritius	71.954	2002
378	Americas	Mexico	74.902	2002
379	Europe	Montenegro	73.981	2002
380	Europe	Netherlands	78.530	2002
381	Oceania	New Zealand	79.110	2002
382	Americas	Nicaragua	70.836	2002
383	Europe	Norway	79.050	2002
384	Asia	Oman	74.193	2002
385	Americas	Panama	74.712	2002
386	Americas	Paraguay	70.755	2002
387	Asia	Philippines	70.303	2002
388	Europe	Poland	74.670	2002
389	Europe	Portugal	77.290	2002
390	Americas	Puerto Rico	77.778	2002
391	Africa	Reunion	75.744	2002
392	Europe	Romania	71.322	2002
393	Asia	Saudi Arabia	71.626	2002

394	Europe	Serbia	73.213	2002
395	Asia	Singapore	78.770	2002
396	Europe	Slovak Republic	73.800	2002
397	Europe	Slovenia	76.660	2002
398	Europe	Spain	79.780	2002
399	Asia	Sri Lanka	70.815	2002
400	Europe	Sweden	80.040	2002
401	Europe	Switzerland	80.620	2002
402	Asia	Syria	73.053	2002
403	Asia	Taiwan	76.990	2002
404	Africa	Tunisia	73.042	2002
405	Europe	Turkey	70.845	2002
406	Europe	United Kingdom	78.471	2002
407	Americas	United States	77.310	2002
408	Americas	Uruguay	75.307	2002
409	Americas	Venezuela	72.766	2002
410	Asia	Vietnam	73.017	2002
411	Asia	West Bank and Gaza	72.370	2002
412	Europe	Albania	76.423	2007
413	Africa	Algeria	72.301	2007
414	Americas	Argentina	75.320	2007
415	Oceania	Australia	81.235	2007
416	Europe	Austria	79.829	2007
417	Asia	Bahrain	75.635	2007
418	Europe	Belgium	79.441	2007
419	Europe	Bosnia and Herzegovina	74.852	2007
420	Americas	Brazil	72.390	2007
421	Europe	Bulgaria	73.005	2007
422	Americas	Canada	80.653	2007
423	Americas	Chile	78.553	2007
424	Asia	China	72.961	2007
425	Americas	Colombia	72.889	2007
426	Americas	Costa Rica	78.782	2007
427	Europe	Croatia	75.748	2007
428	Americas	Cuba	78.273	2007
429	Europe	Czech Republic	76.486	2007
430	Europe	Denmark	78.332	2007
431	Americas	Dominican Republic	72.235	2007
432	Americas	Ecuador	74.994	2007
433	Africa	Egypt	71.338	2007
434	Americas	El Salvador	71.878	2007
435	Europe	Finland	79.313	2007
436	Europe	France	80.657	2007

```
437    Europe              Germany  79.406 2007
438    Europe               Greece  79.483 2007
439  Americas            Guatemala  70.259 2007
440  Americas             Honduras  70.198 2007
441      Asia     Hong Kong, China  82.208 2007
442    Europe              Hungary  73.338 2007
443    Europe              Iceland  81.757 2007
444      Asia            Indonesia  70.650 2007
445      Asia                 Iran  70.964 2007
446    Europe              Ireland  78.885 2007
447      Asia               Israel  80.745 2007
448    Europe                Italy  80.546 2007
449  Americas              Jamaica  72.567 2007
450      Asia                Japan  82.603 2007
451      Asia               Jordan  72.535 2007
452      Asia          Korea, Rep.  78.623 2007
453      Asia               Kuwait  77.588 2007
454      Asia              Lebanon  71.993 2007
455    Africa                Libya  73.952 2007
456      Asia             Malaysia  74.241 2007
457    Africa            Mauritius  72.801 2007
458  Americas               Mexico  76.195 2007
459    Europe           Montenegro  74.543 2007
460    Africa              Morocco  71.164 2007
461    Europe          Netherlands  79.762 2007
462   Oceania          New Zealand  80.204 2007
463  Americas            Nicaragua  72.899 2007
464    Europe               Norway  80.196 2007
465      Asia                 Oman  75.640 2007
466  Americas               Panama  75.537 2007
467  Americas             Paraguay  71.752 2007
468  Americas                 Peru  71.421 2007
469      Asia          Philippines  71.688 2007
470    Europe               Poland  75.563 2007
471    Europe             Portugal  78.098 2007
472  Americas          Puerto Rico  78.746 2007
473    Africa              Reunion  76.442 2007
474    Europe              Romania  72.476 2007
475      Asia         Saudi Arabia  72.777 2007
476    Europe               Serbia  74.002 2007
477      Asia            Singapore  79.972 2007
478    Europe      Slovak Republic  74.663 2007
479    Europe             Slovenia  77.926 2007
```

```
480      Europe                Spain  80.941 2007
481        Asia            Sri Lanka  72.396 2007
482      Europe               Sweden  80.884 2007
483      Europe          Switzerland  81.701 2007
484        Asia                Syria  74.143 2007
485        Asia               Taiwan  78.400 2007
486        Asia             Thailand  70.616 2007
487      Africa              Tunisia  73.923 2007
488      Europe               Turkey  71.777 2007
489      Europe       United Kingdom  79.425 2007
490    Americas        United States  78.242 2007
491    Americas              Uruguay  76.384 2007
492    Americas            Venezuela  73.747 2007
493        Asia              Vietnam  74.249 2007
494        Asia   West Bank and Gaza  73.422 2007
```

- The earliest year that the life expectancy was over 70 was 1952, which was right after WW2. I was surprised that the life expectancy was so high even that early. The country that had the highest life expectancy in 1952 was Iceland, which tracks with the fact that they have the highest life expectancy on average over the whole dataset.

9. Use the functions in dplyr such as filter(), select(), mutate(), etc to explore your data and identify something interesting in your the gapminder data frame. Describe what you identified. (2)

```
gapMinderData %>% dplyr::select(continent, country, gdpPercap, year) %>% filter(year == 20
```

```
   continent             country  gdpPercap year
1     Africa    Congo, Dem. Rep.   241.1659 2002
2     Africa             Burundi   446.4035 2002
3     Africa            Ethiopia   530.0535 2002
4     Africa             Liberia   531.4824 2002
5     Africa       Guinea-Bissau   575.7047 2002
6     Africa               Niger   601.0745 2002
7       Asia             Myanmar   611.0000 2002
8     Africa          Mozambique   633.6179 2002
9     Africa              Gambia   660.5856 2002
10    Africa              Malawi   665.4231 2002
11    Africa            Zimbabwe   672.0386 2002
12    Africa        Sierra Leone   699.4897 2002
```

16

```
13      Asia                 Afghanistan     726.7341 2002
14      Africa  Central African Republic     738.6906 2002
15      Africa                   Eritrea     765.3500 2002
16      Africa                    Rwanda     785.6538 2002
17      Africa                   Somalia     882.0818 2002
18      Africa                      Togo     886.2206 2002
19      Africa                Madagascar     894.6371 2002
20      Asia                    Cambodia     896.2260 2002
21      Africa                  Tanzania     899.0742 2002
22      Africa                    Uganda     927.7210 2002
23      Africa                    Guinea     945.5836 2002
24      Africa                      Mali     951.4098 2002
25      Africa                Burkina Faso  1037.6452 2002
26      Asia                       Nepal    1057.2063 2002
27      Africa                    Zambia    1071.6139 2002
28      Africa                   Comoros    1075.8116 2002
29      Africa                     Ghana    1111.9846 2002
30      Asia                  Bangladesh    1136.3904 2002
31      Africa                      Chad    1156.1819 2002
32      Americas                   Haiti    1270.3649 2002
33      Africa                   Lesotho    1275.1846 2002
34      Africa                     Kenya    1287.5147 2002
35      Africa    Sao Tome and Principe    1353.0924 2002
36      Africa                     Benin    1372.8779 2002
37      Africa                   Senegal    1519.6353 2002
38      Africa                Mauritania    1579.0195 2002
39      Africa                   Nigeria    1615.2864 2002
40      Asia            Korea, Dem. Rep.    1646.7582 2002
41      Africa              Cote d'Ivoire   1648.8008 2002
42      Asia                       India    1746.7695 2002
43      Asia                     Vietnam    1764.4567 2002
44      Africa                  Djibouti    1908.2609 2002
45      Africa                  Cameroon    1934.0114 2002
46      Africa                     Sudan    1993.3983 2002
47      Asia                    Pakistan    2092.7124 2002
48      Asia                    Mongolia    2140.7393 2002
49      Asia                  Yemen, Rep.   2234.8208 2002
50      Americas               Nicaragua    2474.5488 2002
51      Asia                  Philippines   2650.9211 2002
52      Africa                    Angola    2773.2873 2002
53      Asia                   Indonesia    2873.9129 2002
54      Asia                   Sri Lanka    3015.3788 2002
55      Americas                Honduras    3099.7287 2002
```

```
56      Asia                          China  3119.2809 2002
57      Africa                      Morocco  3258.4956 2002
58    Americas                      Bolivia  3413.2627 2002
59      Africa                 Congo, Rep.   3484.0620 2002
60    Americas                     Paraguay  3783.6742 2002
61      Asia                         Jordan  3844.9172 2002
62      Africa                     Namibia   4072.3248 2002
63      Asia                          Syria  4090.9253 2002
64      Africa                    Swaziland  4128.1169 2002
65      Asia                           Iraq  4390.7173 2002
66      Asia           West Bank and Gaza    4515.4876 2002
67    Americas           Dominican Republic  4563.8082 2002
68     Europe                       Albania  4604.2117 2002
69      Africa                        Egypt  4754.6044 2002
70    Americas                    Guatemala  4858.3475 2002
71      Africa                      Algeria  5288.0404 2002
72    Americas                 El Salvador   5351.5687 2002
73      Africa                      Tunisia  5722.8957 2002
74    Americas                     Colombia  5755.2600 2002
75    Americas                      Ecuador  5773.0445 2002
76    Americas                         Peru  5909.0201 2002
77      Asia                       Thailand  5913.1875 2002
78     Europe    Bosnia and Herzegovina     6018.9752 2002
79      Africa                       Reunion 6316.1652 2002
80    Americas                         Cuba  6340.6467 2002
81     Europe                        Turkey  6508.0857 2002
82     Europe                    Montenegro  6557.1943 2002
83    Americas                      Jamaica  6994.7749 2002
84     Europe                        Serbia  7236.0753 2002
85    Americas                       Panama  7356.0319 2002
86     Europe                      Bulgaria  7696.7777 2002
87      Africa           Equatorial Guinea   7703.4959 2002
88      Africa                South Africa   7710.9464 2002
89    Americas                   Costa Rica  7723.4472 2002
90    Americas                      Uruguay  7727.0020 2002
91     Europe                      Romania   7885.3601 2002
92    Americas                       Brazil  8131.2128 2002
93    Americas                    Venezuela  8605.0478 2002
94    Americas                    Argentina  8797.6407 2002
95      Africa                    Mauritius  9021.8159 2002
96      Asia                          Iran   9240.7620 2002
97      Asia                       Lebanon   9313.9388 2002
98      Africa                        Libya  9534.6775 2002
```

```
99       Asia                  Malaysia 10206.9779 2002
100   Americas                   Mexico 10742.4405 2002
101   Americas                    Chile 10778.7838 2002
102     Africa                 Botswana 11003.6051 2002
103   Americas      Trinidad and Tobago 11460.6002 2002
104     Europe                  Croatia 11628.3890 2002
105     Europe                   Poland 12002.2391 2002
106     Africa                    Gabon 12521.7139 2002
107     Europe          Slovak Republic 13638.7784 2002
108     Europe                  Hungary 14843.9356 2002
109     Europe           Czech Republic 17596.2102 2002
110   Americas              Puerto Rico 18855.6062 2002
111       Asia             Saudi Arabia 19014.5412 2002
112       Asia              Korea, Rep. 19233.9882 2002
113       Asia                     Oman 19774.8369 2002
114     Europe                 Portugal 19970.9079 2002
115     Europe                 Slovenia 20660.0194 2002
116       Asia                   Israel 21905.5951 2002
117     Europe                   Greece 22514.2548 2002
118    Oceania              New Zealand 23189.8014 2002
119       Asia                   Taiwan 23235.4233 2002
120       Asia                  Bahrain 23403.5593 2002
121     Europe                    Spain 24835.4717 2002
122     Europe                    Italy 27968.0982 2002
123     Europe                  Finland 28204.5906 2002
124       Asia                    Japan 28604.5919 2002
125     Europe                   France 28926.0323 2002
126     Europe                   Sweden 29341.6309 2002
127     Europe           United Kingdom 29478.9992 2002
128     Europe                  Germany 30035.8020 2002
129       Asia         Hong Kong, China 30209.0152 2002
130     Europe                  Belgium 30485.8838 2002
131    Oceania                Australia 30687.7547 2002
132     Europe                  Iceland 31163.2020 2002
133     Europe                  Denmark 32166.5001 2002
134     Europe                  Austria 32417.6077 2002
135   Americas                   Canada 33328.9651 2002
136     Europe              Netherlands 33724.7578 2002
137     Europe                  Ireland 34077.0494 2002
138     Europe              Switzerland 34480.9577 2002
139       Asia                   Kuwait 35110.1057 2002
140       Asia                Singapore 36023.1054 2002
141   Americas            United States 39097.0995 2002
```
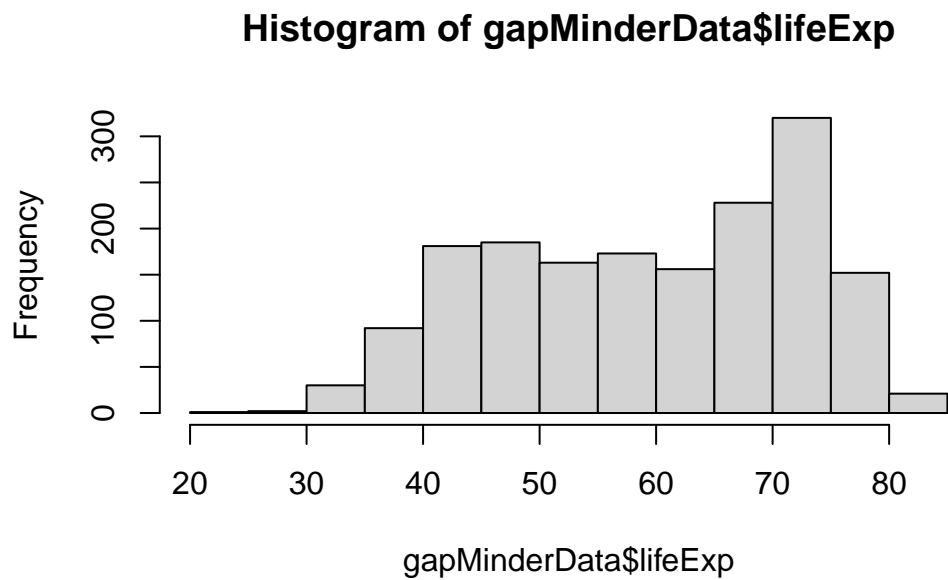
```
142       Europe                           Norway 44683.9753 2002
```

- In 2002, the country with the highest GDP was Norway and the country with the lowest GDP was the DRC. The difference between the countries is about 44442.81. This is a very large amount.
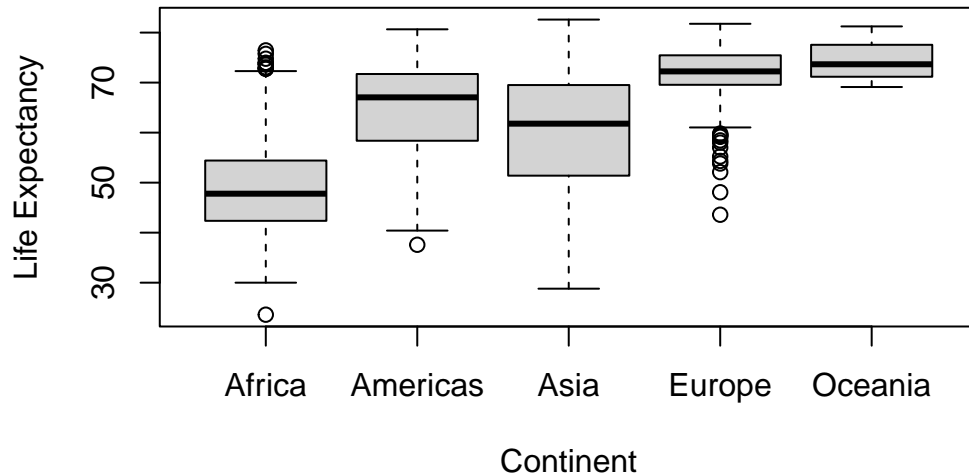
10. Make a histogram of Life Expectancy with base R plotting functions

```
hist(gapMinderData$lifeExp)
```

Histogram of gapMinderData$lifeExp

11. Make a plot of your choosing

```
boxplot(gapMinderData$lifeExp ~ gapMinderData$continent, ylab = "Life Expectancy", xlab =
```

12. How do R and Python index differently

- R indexes from 1 while Python indexes from 0.

13. What does the function c() in R and the function concatenate() do?

- c() in R combines values into a vector and returns a vector in a 1D array
- concatenate() in R combines two string/character vectors into one vector

14. What is the most important thing you should consider when starting a data analysis project for someone?

- Have the question that the person wants answered clear so that the scope of the project is clear, the observations can be more relevant, and the final users of the analysis can understand the analysis completely.

15. What is the difference between discrete and continuous data?

- Discrete data is represented as exact values that can be counted (integer values)
- Continuous data contains values that are measured over a time interval (can contains complex numbers)