

# Assignment 2: Gapminder and ‘dplyr’

Hannah McSorley

21/09/2019

## Homework Assignment 2

The Objective of this assignment is the practice using ggplot2 and dplyr to explore and visualize the gapminder dataset.

```
suppressPackageStartupMessages(library(gapminder))
```

```
## Warning: package 'gapminder' was built under R version 3.5.3
```

```
suppressPackageStartupMessages(library(tidyverse)) # included package 'dplyr'
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## Warning: package 'readr' was built under R version 3.5.3
```

```
## Warning: package 'purrr' was built under R version 3.5.3
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
## Warning: package 'stringr' was built under R version 3.5.3
```

```
## Warning: package 'forcats' was built under R version 3.5.3
```

```
suppressPackageStartupMessages(library(tsibble))
```

```
## Warning: package 'tsibble' was built under R version 3.5.3
```

```
suppressPackageStartupMessages(library(knitr)) # tidy tables
```

```
suppressPackageStartupMessages(library(DT)) # tidy dataframe tables for HTML
```

## ‘dplyr’ function focus

Within the ‘dplyr’ library, important functions to employ are:

- *filter()*: a tidy way to subset *in-situ*
- *%>%*: Piping to simplify multiple actions on a dataset

## 1.1 filter()

Subset the ‘gapminder’ data to three countries in the 1970’s using the ‘filter()’ function.

```
# filter for the 1970's for three countries
gapminder %>%
  dplyr::filter(between(year, 1970, 1979),
                country == "Puerto Rico" | country == "Yemen, Rep." | country == "Zimbabwe")

## # A tibble: 6 x 6
##   country      continent year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int>   <dbl>   <int>   <dbl>
## 1 Puerto Rico Americas  1972   72.2  2847132   9123.
## 2 Puerto Rico Americas  1977   73.4  3080828   9771.
## 3 Yemen, Rep. Asia      1972   39.8  7407075   1265.
## 4 Yemen, Rep. Asia      1977   44.2  8403990   1830.
## 5 Zimbabwe     Africa   1972   55.6  5861135    799.
## 6 Zimbabwe     Africa   1977   57.7  6642107    686.

# countries were selected randomly using the function 'sample_n()'
# sample_n() selects n number of rows randomly
# gapminder %>% filter(between(year, 1970, 1979)) %>% sample_n(size = 3)

# I want to use this filtered dataset again, so save it as a new object
seventies_three <- gapminder %>%
  dplyr::filter(between(year, 1970, 1979),
                country == "Puerto Rico" | country == "Yemen, Rep." | country == "Zimbabwe")
```

## 1.2 pipe

Use the pipe operator %>% to select “country” and “gdpPercap” from your filtered dataset in 1.1.

```
# subset the dataset created in 1.1
seventies_three %>%
  select(country, gdpPercap)
```

```
## # A tibble: 6 x 2
##   country      gdpPercap
##   <fct>        <dbl>
## 1 Puerto Rico   9123.
## 2 Puerto Rico   9771.
## 3 Yemen, Rep.   1265.
## 4 Yemen, Rep.   1830.
## 5 Zimbabwe      799.
## 6 Zimbabwe      686.
```

## 1.3 Advanced Filtering

Filter gapminder to all entries that have experienced a drop in life expectancy. Be sure to include a new variable that’s the increase in life expectancy in your tibble. Hint: you might find the lag() or diff() functions useful.

```
# calculate the difference in life expectancy over time for each country
# filter for those with a drop in life expectancy
```

```
gapminder %>%
  arrange(country, year) %>%
  mutate(change_lifeExp = tsibble::difference(lifeExp)) %>%
  filter(change_lifeExp < 0) %>%
  head() %>%
  kable(digits = 2)
```

country	continent	year	lifeExp	pop	gdpPercap	change_lifeExp
Albania	Europe	1992	71.58	3326498	2497.44	-0.42
Algeria	Africa	1952	43.08	9279525	2449.01	-33.35
Angola	Africa	1952	30.02	4232095	3520.61	-42.29
Angola	Africa	1987	39.91	7874230	2430.21	-0.04
Australia	Oceania	1952	69.12	8691212	10039.60	-6.20
Austria	Europe	1952	66.80	6927772	6137.08	-14.44

```
# unclear instruction about including a new variable for the increase in life expectancy...
# when we're filtering for the drop in life expectancy
```

```
# maybe something that identifies whether there was an increase or decrease?
```

```
gapminder %>%
  arrange(country, year) %>%
  drop_na() %>%
  mutate(change_lifeExp = tsibble::difference(lifeExp),
         incr_lifeExp = (change_lifeExp > 0)) %>%
  head() %>%
  kable(digits = 2)
```

country	continent	year	lifeExp	pop	gdpPercap	change_lifeExp	incr_lifeExp
Afghanistan	Asia	1952	28.80	8425333	779.45	NA	NA
Afghanistan	Asia	1957	30.33	9240934	820.85	1.53	TRUE
Afghanistan	Asia	1962	32.00	10267083	853.10	1.66	TRUE
Afghanistan	Asia	1967	34.02	11537966	836.20	2.02	TRUE
Afghanistan	Asia	1972	36.09	13079460	739.98	2.07	TRUE
Afghanistan	Asia	1977	38.44	14880372	786.11	2.35	TRUE

```
# provides TRUE or FALSE indicator of whether the lifeExp increased for that entry,
# relative to the previous data
```

```
# filter the above to include only entries that have had a drop in life expectancy
```

```
gapminder %>%
  arrange(country, year) %>%
  drop_na() %>%
  mutate(change_lifeExp = tsibble::difference(lifeExp),
         incr_lifeExp = (change_lifeExp > 0)) %>%
  filter(incr_lifeExp == FALSE) %>%
  select(country:change_lifeExp) %>%
  kable(digits = 2, caption = "drop in life expectancy")
```

Table 3: drop in life expectancy

country	continent	year	lifeExp	pop	gdpPercap	change_lifeExp
Albania	Europe	1992	71.58	3326498	2497.44	-0.42
Algeria	Africa	1952	43.08	9279525	2449.01	-33.35
Angola	Africa	1952	30.02	4232095	3520.61	-42.29
Angola	Africa	1987	39.91	7874230	2430.21	-0.04
Australia	Oceania	1952	69.12	8691212	10039.60	-6.20
Austria	Europe	1952	66.80	6927772	6137.08	-14.44
Bahrain	Asia	1952	50.94	120447	9867.08	-28.89
Bangladesh	Asia	1952	37.48	46886859	684.24	-38.15
Benin	Africa	1952	38.22	1738315	1062.75	-41.22
Benin	Africa	2002	54.41	7026113	1372.88	-0.37
Bolivia	Americas	1952	40.41	2883315	2677.33	-16.31
Bosnia and Herzegovina	Europe	1952	53.82	2791000	973.53	-11.73
Botswana	Africa	1952	47.62	442308	851.24	-27.23
Botswana	Africa	1992	62.74	1342614	7954.11	-0.88
Botswana	Africa	1997	52.56	1536536	8647.14	-10.19
Botswana	Africa	2002	46.63	1630347	11003.61	-5.92
Bulgaria	Europe	1952	59.60	7274900	2444.29	-12.79
Bulgaria	Europe	1977	70.81	8797022	7612.24	-0.09
Bulgaria	Europe	1992	71.19	8658506	6302.62	-0.15
Bulgaria	Europe	1997	70.32	8066057	5970.39	-0.87
Burkina Faso	Africa	1952	31.98	4469979	543.26	-41.03
Burundi	Africa	1952	39.03	2445618	339.30	-13.26
Burundi	Africa	1992	44.74	5809236	631.70	-3.48
Cambodia	Asia	1952	39.42	4693836	368.47	-10.16
Cambodia	Asia	1972	40.32	7450606	421.62	-5.10
Cambodia	Asia	1977	31.22	6978607	524.97	-9.10
Cameroon	Africa	1952	38.52	5009067	1172.67	-21.20
Cameroon	Africa	1992	54.31	12467171	1793.16	-0.67
Cameroon	Africa	1997	52.20	14195809	1694.34	-2.12
Cameroon	Africa	2002	49.86	15929988	1934.01	-2.34
Central African Republic	Africa	1952	35.46	1291695	1071.31	-45.19
Central African Republic	Africa	1992	49.40	3265124	747.91	-1.09
Central African Republic	Africa	1997	46.07	3696513	740.51	-3.33
Central African Republic	Africa	2002	43.31	4048013	738.69	-2.76
Chad	Africa	1952	38.09	2682462	1178.67	-6.65
Chad	Africa	1997	51.57	7562011	1004.96	-0.15
Chad	Africa	2002	50.52	8835739	1156.18	-1.05
China	Asia	1952	44.00	556263527	400.45	-34.55
China	Asia	1962	44.50	665770000	487.67	-6.05
Colombia	Americas	1952	50.64	12350771	2144.12	-22.32
Comoros	Africa	1952	40.72	153936	1102.99	-32.17
Congo, Dem. Rep.	Africa	1952	39.14	14100005	780.54	-26.01
Congo, Dem. Rep.	Africa	1982	47.78	30646495	673.75	-0.02
Congo, Dem. Rep.	Africa	1987	47.41	35481645	672.77	-0.37
Congo, Dem. Rep.	Africa	1992	45.55	41672143	457.72	-1.86
Congo, Dem. Rep.	Africa	1997	42.59	47798986	312.19	-2.96
Congo, Rep.	Africa	1952	42.11	854885	2125.62	-4.35
Congo, Rep.	Africa	1992	56.43	2409073	4016.24	-1.04
Congo, Rep.	Africa	1997	52.96	2800947	3484.16	-3.47
Cote d'Ivoire	Africa	1952	40.48	2977019	1388.59	-38.30

country	continent	year	lifeExp	pop	gdpPercap	change_lifeExp
Cote d'Ivoire	Africa	1992	52.04	12772596	1648.07	-2.61
Cote d'Ivoire	Africa	1997	47.99	14625967	1786.27	-4.05
Cote d'Ivoire	Africa	2002	46.83	16252726	1648.80	-1.16
Croatia	Europe	1982	70.46	4413368	13221.82	-0.18
Cuba	Americas	1952	59.42	6007797	5586.54	-16.33
Czech Republic	Europe	1952	66.87	9125183	6876.14	-11.40
Czech Republic	Europe	1972	70.29	9862158	13108.45	-0.09
Denmark	Europe	1952	70.78	4334000	9692.39	-5.71
Denmark	Europe	1982	74.63	5117810	21688.04	-0.06
Djibouti	Africa	1952	34.81	63149	2669.53	-43.52
Dominican Republic	Americas	1952	45.93	2491346	1397.72	-8.86
Ecuador	Americas	1952	48.36	3548753	3522.11	-23.88
Egypt	Africa	1952	41.89	22223309	1418.82	-33.10
El Salvador	Americas	1952	45.26	2042865	3048.30	-26.08
El Salvador	Americas	1977	56.70	4282586	5138.92	-1.51
El Salvador	Americas	1982	56.60	4474873	4098.34	-0.09
Equatorial Guinea	Africa	1952	34.48	216964	375.64	-37.40
Eritrea	Africa	1952	35.93	1438760	328.94	-15.65
Eritrea	Africa	1982	43.89	2637297	524.88	-0.64
Ethiopia	Africa	1952	34.08	20860941	362.15	-23.96
France	Europe	1952	67.41	42459667	7029.81	-11.90
Gabon	Africa	1952	37.00	420702	4293.48	-43.65
Gabon	Africa	1997	60.46	1126189	14722.84	-0.91
Gabon	Africa	2002	56.76	1299304	12521.71	-3.70
Gabon	Africa	2007	56.73	1454867	13206.48	-0.03
Gambia	Africa	1952	30.00	284320	485.23	-26.73
Ghana	Africa	1952	43.15	5581001	911.30	-36.26
Ghana	Africa	2002	58.45	20550751	1111.98	-0.10
Guatemala	Americas	1952	42.02	3146381	2428.24	-37.46
Guinea	Africa	1952	33.61	2664249	510.20	-36.65
Guinea-Bissau	Africa	1952	32.50	580653	299.85	-23.51
Haiti	Americas	1952	37.58	3201488	1840.37	-8.81
Honduras	Americas	1952	41.91	1517453	2194.93	-19.00
Hong Kong, China	Asia	1952	60.96	2125900	3054.42	-9.24
Hungary	Europe	1952	64.03	9504000	5263.67	-18.18
Hungary	Europe	1982	69.39	10705535	12545.99	-0.56
Hungary	Europe	1992	69.17	10348684	10535.63	-0.41
Iceland	Europe	1952	72.49	147962	7267.69	-0.85
India	Asia	1952	37.37	372000000	546.57	-44.38
Indonesia	Asia	1952	37.47	82052000	749.68	-27.23
Iran	Asia	1952	44.87	17272000	3035.33	-25.78
Iraq	Asia	1952	45.32	5441766	4129.77	-25.64
Iraq	Asia	1992	59.46	17861905	3745.64	-5.58
Iraq	Asia	1997	58.81	20775703	3076.24	-0.65
Iraq	Asia	2002	57.05	24001816	4390.72	-1.77
Israel	Asia	1952	65.39	1620914	4086.52	-13.50
Italy	Europe	1952	65.94	47666000	4931.40	-14.81
Jamaica	Americas	1952	58.53	1426095	2898.53	-22.02
Jamaica	Americas	1992	71.77	2378618	7404.92	0.00
Jamaica	Americas	2002	72.05	2664659	6994.77	-0.22
Japan	Asia	1952	63.03	86459025	3216.96	-9.54
Jordan	Asia	1952	43.16	607914	1546.91	-39.44

country	continent	year	lifeExp	pop	gdpPercap	change_lifeExp
Kenya	Africa	1952	42.27	6464046	853.54	-30.26
Kenya	Africa	1992	59.28	25020539	1341.92	-0.05
Kenya	Africa	1997	54.41	28263827	1360.49	-4.88
Kenya	Africa	2002	50.99	31386842	1287.51	-3.41
Korea, Dem. Rep.	Asia	1952	50.06	8865488	1088.28	-4.05
Korea, Dem. Rep.	Asia	1992	69.98	20711375	3726.06	-0.67
Korea, Dem. Rep.	Asia	1997	67.73	21585105	1690.76	-2.25
Korea, Dem. Rep.	Asia	2002	66.66	22215365	1646.76	-1.06
Korea, Rep.	Asia	1952	47.45	20947571	1030.59	-19.84
Kuwait	Asia	1952	55.56	160000	108382.35	-23.06
Lebanon	Asia	1952	55.93	1439529	4834.80	-21.66
Lesotho	Africa	1952	42.14	748747	298.85	-29.85
Lesotho	Africa	1997	55.56	1982823	1186.15	-4.13
Lesotho	Africa	2002	44.59	2046772	1275.18	-10.96
Lesotho	Africa	2007	42.59	2012649	1569.33	-2.00
Liberia	Africa	1952	38.48	863308	575.57	-4.11
Liberia	Africa	1992	40.80	1912974	636.62	-5.23
Libya	Africa	1952	42.72	1019729	2387.55	-2.95
Madagascar	Africa	1952	36.68	4762912	1443.01	-37.27
Malawi	Africa	1952	36.26	2917802	369.17	-23.19
Malawi	Africa	1997	47.49	10419991	692.28	-1.93
Malawi	Africa	2002	45.01	11824495	665.42	-2.49
Mali	Africa	1952	33.69	3838168	452.34	-40.56
Mauritania	Africa	1952	40.54	1022556	743.12	-13.92
Mauritius	Africa	1952	50.99	516556	1967.96	-13.18
Mexico	Americas	1952	50.79	30144317	3478.13	-22.01
Mongolia	Asia	1952	42.24	800663	786.57	-33.95
Montenegro	Europe	1952	59.16	413834	2647.59	-7.64
Montenegro	Europe	2002	73.98	720230	6557.19	-1.46
Morocco	Africa	1952	42.87	9939217	1688.20	-31.67
Mozambique	Africa	1952	31.29	6446316	468.53	-39.88
Mozambique	Africa	2002	44.03	18473780	633.62	-2.32
Mozambique	Africa	2007	42.08	19951656	823.69	-1.94
Myanmar	Asia	1952	36.32	20092996	331.00	-5.76
Myanmar	Asia	2002	59.91	45598081	611.00	-0.42
Namibia	Africa	1952	41.73	485831	2423.78	-20.34
Namibia	Africa	1997	58.91	1774766	3899.52	-3.09
Namibia	Africa	2002	51.48	1972153	4072.32	-7.43
Nepal	Asia	1952	36.16	9182536	545.87	-16.75
Netherlands	Europe	1972	73.75	13329874	18794.75	-0.07
New Zealand	Oceania	1952	69.39	1994794	10556.58	-10.37
Nicaragua	Americas	1952	42.31	1165790	3112.36	-37.89
Niger	Africa	1952	37.44	3379468	761.88	-35.45
Nigeria	Africa	1952	36.32	33119096	1077.28	-20.54
Nigeria	Africa	1997	47.46	106207839	1624.94	-0.01
Nigeria	Africa	2002	46.61	119901274	1615.29	-0.86
Norway	Europe	1987	75.89	4186147	31540.97	-0.08
Oman	Asia	1952	37.58	507833	1828.23	-42.62
Pakistan	Asia	1952	43.44	41346560	684.60	-32.20
Panama	Americas	1952	55.19	940080	2480.38	-10.29
Paraguay	Americas	1952	62.65	1555876	1952.31	-12.89
Peru	Americas	1952	43.90	8025700	3758.52	-27.85

country	continent	year	lifeExp	pop	gdpPercap	change_lifeExp
Philippines	Asia	1952	47.75	22438691	1272.88	-23.67
Poland	Europe	1952	61.31	25730551	4029.33	-10.38
Poland	Europe	1977	70.67	34621254	9508.14	-0.18
Poland	Europe	1987	70.98	37740710	9082.35	-0.34
Portugal	Europe	1952	59.82	8526050	3068.32	-15.74
Puerto Rico	Americas	1952	64.28	2227000	3081.96	-13.82
Puerto Rico	Americas	1992	73.91	3585176	14641.59	-0.72
Reunion	Africa	1952	52.72	257700	2718.89	-26.02
Romania	Europe	1952	61.05	16630000	3144.61	-15.39
Romania	Europe	1967	66.80	19284814	6470.87	0.00
Romania	Europe	1987	69.53	22686371	9696.27	-0.13
Romania	Europe	1992	69.36	22797027	6598.41	-0.17
Rwanda	Africa	1952	40.00	2534927	493.32	-32.48
Rwanda	Africa	1987	44.02	6349365	847.99	-2.20
Rwanda	Africa	1992	23.60	7290203	737.07	-20.42
Saudi Arabia	Asia	1952	39.88	4005677	6459.55	-25.65
Senegal	Africa	1952	37.28	2755589	1450.36	-35.50
Serbia	Europe	1952	58.00	6860147	3581.46	-5.07
Serbia	Europe	1982	70.16	9032824	15181.09	-0.14
Sierra Leone	Africa	1952	30.33	2143249	879.79	-43.67
Sierra Leone	Africa	1992	38.33	4260884	1068.70	-1.67
Slovak Republic	Europe	1952	64.36	3558137	5074.66	-15.61
Slovak Republic	Europe	1972	70.35	4593433	9674.17	-0.63
Slovenia	Europe	1952	65.57	1489518	4215.04	-9.09
Somalia	Africa	1952	32.98	2526994	1135.75	-44.95
Somalia	Africa	1992	39.66	6099799	926.96	-4.84
South Africa	Africa	1952	45.01	14264935	4725.30	-3.15
South Africa	Africa	1997	60.24	42835005	7479.19	-1.65
South Africa	Africa	2002	53.37	44433622	7710.95	-6.87
South Africa	Africa	2007	49.34	43997828	9269.66	-4.03
Sri Lanka	Asia	1952	57.59	7982342	1083.53	-23.35
Sudan	Africa	1952	38.63	8504667	1615.99	-33.76
Swaziland	Africa	1952	41.41	290243	1148.38	-17.15
Swaziland	Africa	1997	54.29	1054486	3876.77	-4.18
Swaziland	Africa	2002	43.87	1130269	4128.12	-10.42
Swaziland	Africa	2007	39.61	1133066	4513.48	-4.26
Switzerland	Europe	1952	69.62	4815000	14734.23	-11.26
Syria	Asia	1952	45.88	3661549	1643.49	-35.82
Taiwan	Asia	1952	58.50	8550362	1206.95	-15.64
Tanzania	Africa	1952	41.22	8322925	716.65	-37.19
Tanzania	Africa	1992	50.44	26605473	825.68	-1.09
Tanzania	Africa	1997	48.47	30686889	789.19	-1.97
Thailand	Asia	1952	50.85	21289402	757.80	-1.67
Togo	Africa	1952	38.60	1219113	859.81	-32.02
Togo	Africa	2002	57.56	4977378	886.22	-0.83
Trinidad and Tobago	Americas	1997	69.47	1138101	8792.57	-0.40
Trinidad and Tobago	Americas	2002	68.98	1101832	11460.60	-0.49
Tunisia	Africa	1952	44.60	3647735	1468.48	-25.22
Turkey	Europe	1952	43.59	22235677	1969.10	-30.34
Uganda	Africa	1952	39.98	5824797	734.75	-31.80
Uganda	Africa	1977	50.35	11457758	843.73	-0.67
Uganda	Africa	1982	49.85	12939400	682.27	-0.50

country	continent	year	lifeExp	pop	gdpPercap	change_lifeExp
Uganda	Africa	1992	48.83	18252190	644.17	-2.68
Uganda	Africa	1997	44.58	21210254	816.56	-4.25
United States	Americas	1952	68.44	157553000	13990.48	-10.98
Uruguay	Americas	1952	66.07	2252965	5716.77	-12.17
Venezuela	Americas	1952	55.09	5439568	7689.80	-21.30
Vietnam	Asia	1952	40.41	26246839	605.07	-33.34
West Bank and Gaza	Asia	1952	43.16	1030585	1515.59	-31.09
Yemen, Rep.	Asia	1952	32.55	4963829	781.72	-40.87
Zambia	Africa	1952	42.04	2672000	1147.39	-20.66
Zambia	Africa	1987	50.82	7272406	1213.32	-1.00
Zambia	Africa	1992	46.10	8381163	1210.88	-4.72
Zambia	Africa	1997	40.24	9417789	1071.35	-5.86
Zambia	Africa	2002	39.19	10595811	1071.61	-1.05
Zimbabwe	Africa	1992	60.38	10704340	693.42	-1.97
Zimbabwe	Africa	1997	46.81	11404948	792.45	-13.57
Zimbabwe	Africa	2002	39.99	11926563	672.04	-6.82

## 1.4 Advanced Filtering

Filter gapminder so that it shows the max GDP per capita experienced by each country. Hint: you might find the `max()` function useful here.

```
# create a variable for the maximum GDP per capita in each country
gapminder %>%
  group_by(country) %>%
  mutate(max_gdp = max(gdpPercap)) %>%
  filter(gdpPercap == max_gdp) %>%
  head() %>%
  kable(digits = 2)
```

country	continent	year	lifeExp	pop	gdpPercap	max_gdp
Afghanistan	Asia	1982	39.85	12881816	978.01	978.01
Albania	Europe	2007	76.42	3600523	5937.03	5937.03
Algeria	Africa	2007	72.30	33333216	6223.37	6223.37
Angola	Africa	1967	35.98	5247469	5522.78	5522.78
Argentina	Americas	2007	75.32	40301927	12779.38	12779.38
Australia	Oceania	2007	81.23	20434176	34435.37	34435.37
## 1.4 Alternativa						
<b>OR</b>						

Filter gapminder to contain six rows: the rows with the three largest GDP per capita, and the rows with the three smallest GDP per capita. Be sure to not create any intermediate objects when doing this (with, for example, the assignment operator). Hint: you might find the `sort()` function useful, or perhaps even the `dplyr::slice()` function.

```
# use above code to add max gdpPercap
# filter to 3 highest max_gdp and 3 lowest max_gdp
```

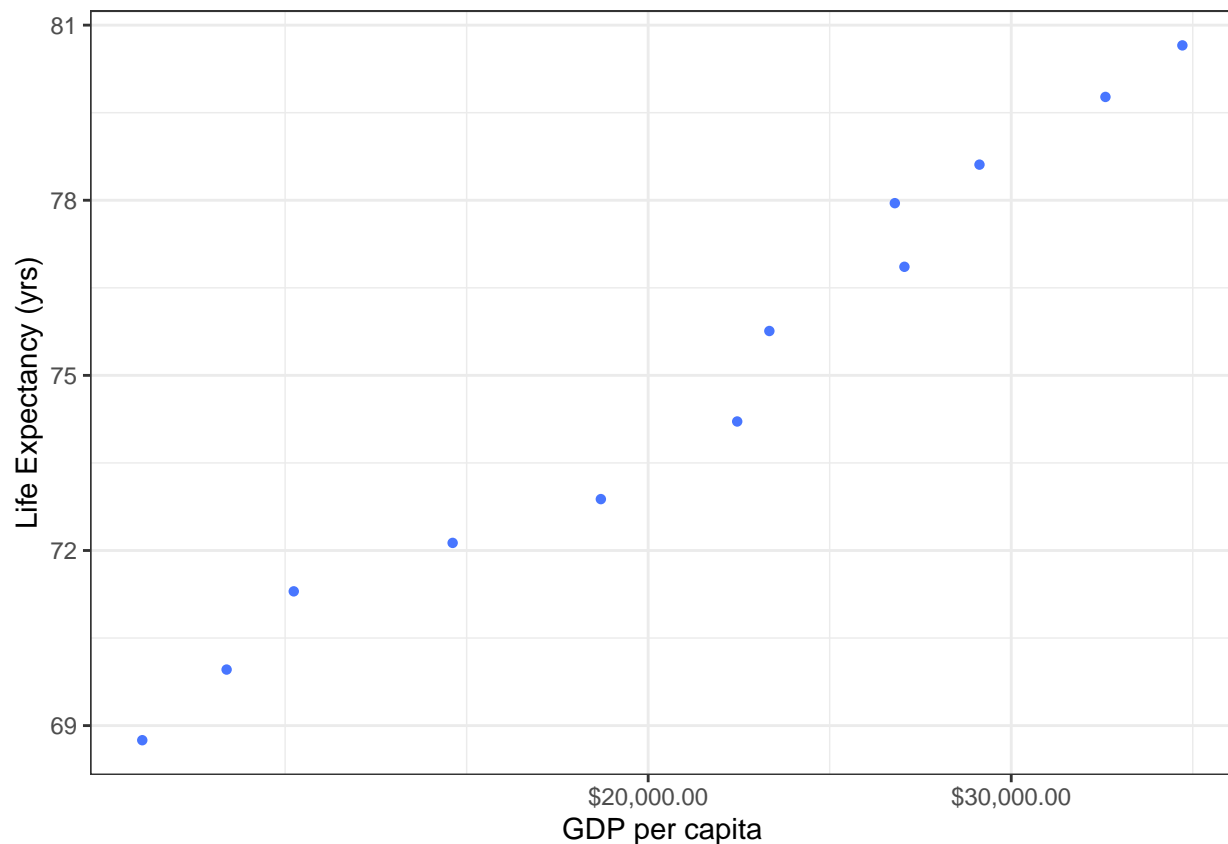


```
# NOPE -- didn't get it this time.
# gapminder %>%
#   group_by(country) %>%
#   mutate(max_gdp = max(gdpPercap)) %>%
#   filter(gdpPercap == max_gdp) %>%
#   arrange(max_gdp) %>%
#   filter(between(row_number(), 1, 3) & between(row_number(), n(), (n()-3)))
```

## 1.5

Produce a scatterplot of Canada's life expectancy vs. GDP per capita using ggplot2, without defining a new variable. That is, after filtering the gapminder data set, pipe it directly into the ggplot() function. Ensure GDP per capita is on a log scale.

```
# life expectancy vs. GDP per capita
# do not create a new variable, pipe filters directly into plot
gapminder %>%
  filter(country == "Canada") %>%
  ggplot(aes(x = gdpPercap, y = lifeExp)) +
  geom_point(colour = "royalblue1", shape = 16) +
  scale_x_log10("GDP per capita", labels = scales::dollar_format()) +
  ylab("Life Expectancy (yrs)") +
  theme_bw()
```



## Exercise 2

Explore individual variables in *dplyr*. Pick one **categorical variable** and one **quantitative variable** to explore. Answer the following questions in whichever way you think is appropriate, using *dplyr*:

-1. What are possible values (or range, whichever is appropriate) of each variable? -2. What values are typical? What's the spread? What's the distribution? Etc., tailored to the variable at hand. \* Feel free to use summary stats, tables, figures.

### 1 – categorical variable: continent (and countries in each)

```
## What are possible values (or range, whichever is appropriate) of each variable
# ---
# how many continents are there in the gapminder dataset?
str(gapminder$continent)
```

```
## Factor w/ 5 levels "Africa","Americas",...: 3 3 3 3 3 3 3 3 3 3 ...
```

```
# what are they?
unique(gapminder$continent)
```

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

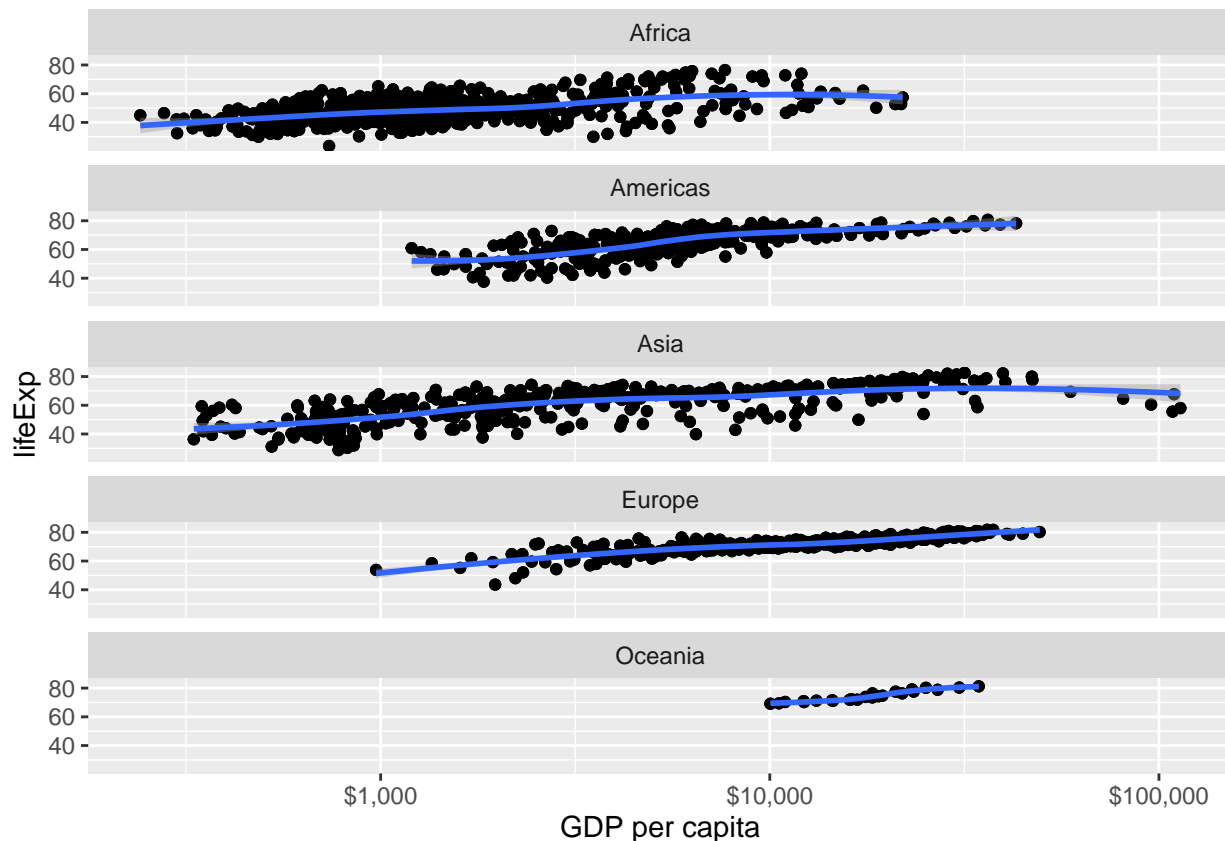
```
# how many countries are in each continent
gapminder %>%
  group_by(continent) %>%
  summarize(countries = n()) %>%
  kable(caption = "number of countries by continent, 'gapminder'",
        col.names = c("Continent", "Country Count"))
```

Table 5: number of countries by continent, 'gapminder'

Continent	Country Count
Africa	624
Americas	300
Asia	396
Europe	360
Oceania	24

```
# plot continents with their GDP per capita and life expectancy
# x axis on a log scale
# get fancy with facet wrapping (multiple panels)
gapminder %>%
  ggplot(aes(x = gdpPercap, y = lifeExp)) +
  geom_point() +
  facet_wrap(~continent) +
  geom_point()
```

```
geom_smooth(method = "loess") +
facet_wrap(~continent, ncol=1) +
scale_x_log10("GDP per capita", labels = scales::dollar_format())
```



There are five continents in the gapminder dataset. Africa has the most countries, followed by America (North, Central, South), then Asia and Europe, and Oceania has the fewest countries.

## 2 – quantitative variable: GDP

*# What values are typical? What's the spread? What's the distribution? Etc.,  
# tailored to the variable at hand.*

```
# check the 'summary' stats
gapminder %>%
  group_by(continent) %>%
  summary()
```

```
##      country      continent      year      lifeExp
## Afghanistan: 12 Africa :624 Min. :1952 Min. :23.60
## Albania : 12 Americas:300 1st Qu.:1966 1st Qu.:48.20
## Algeria : 12 Asia :396 Median :1980 Median :60.71
## Angola : 12 Europe :360 Mean :1980 Mean :59.47
## Argentina : 12 Oceania : 24 3rd Qu.:1993 3rd Qu.:70.85
## Australia : 12 Max. :2007 Max. :82.60
## (Other) :1632
##      pop      gdpPercap
```

```
## Min.      :6.001e+04   Min.      : 241.2
## 1st Qu.:2.794e+06   1st Qu.: 1202.1
## Median :7.024e+06   Median : 3531.8
## Mean    :2.960e+07   Mean    : 7215.3
## 3rd Qu.:1.959e+07   3rd Qu.: 9325.5
## Max.    :1.319e+09   Max.    :113523.1
##
```

```
# long way
# by continent,
# what's the mean, relative standard deviation, mode, min and max GDP per capita?
gapminder %>%
  group_by(continent) %>%
  summarize(mean_gdpp = mean(gdpPercap),
            sd_gdpp = sd(gdpPercap),
            rsd_gdpp = (sd(gdpPercap) / mean_gdpp) * 100,
            med_gdpp = median(gdpPercap),
            min_gdpp = min(gdpPercap),
            max_gdpp = max(gdpPercap)) %>%
  kable(digits = 2, caption = "GDP per capita by continent")
```

Table 6: GDP per capita by continent

continent	mean_gdpp	sd_gdpp	rsd_gdpp	med_gdpp	min_gdpp	max_gdpp
Africa	2193.75	2827.93	128.91	1192.14	241.17	21951.21
Americas	7136.11	6396.76	89.64	5465.51	1201.64	42951.65
Asia	7902.15	14045.37	177.74	2646.79	331.00	113523.13
Europe	14469.48	9355.21	64.65	12081.75	973.53	49357.19
Oceania	18621.61	6358.98	34.15	17983.30	10039.60	34435.37

```
# what if you first multiple by population to get GDP?
gapminder %>%
  mutate(gdpBill = round(gdpPercap*pop/10^9, digits = 2)) %>%
  group_by(continent) %>%
  summarize(mean_gdp = mean(gdpBill),
            sd_gdp = sd(gdpBill),
            rsd_gdp = (sd(gdpBill) / mean_gdp) * 100,
            med_gdp = median(gdpBill),
            min_gdp = min(gdpBill),
            max_gdp = max(gdpBill)) %>%
  kable(digits = 2, caption = "GDP by continent")
```

Table 7: GDP by continent

continent	mean_gdp	sd_gdp	rsd_gdp	med_gdp	min_gdp	max_gdp
Africa	20.90	50.25	240.40	5.33	0.05	447.97
Americas	379.26	1427.43	376.37	29.26	2.00	12934.46
Asia	227.23	628.31	276.51	39.65	0.63	6539.50
Europe	269.44	439.23	163.01	112.43	1.08	2650.87
Oceania	188.19	193.86	103.01	97.14	21.06	703.66

```

# how about summarise GDP per capita and population?
gapminder %>%
  group_by(continent) %>%
  select(gdpPercap, pop, continent) %>%
  summarise(mean_gdp = mean(gdpPercap),
            sd_gdp = sd(gdpPercap),
            rsd_gdp = (sd(gdpPercap) / mean_gdp) * 100,
            mean_pop = mean(pop),
            sd_pop = sd(pop),
            rsd_pop = (sd(pop) / mean_pop) * 100) %>%
  kable(digits = 2, caption = "GDP per capita and population")

```

Table 8: GDP per capita and population

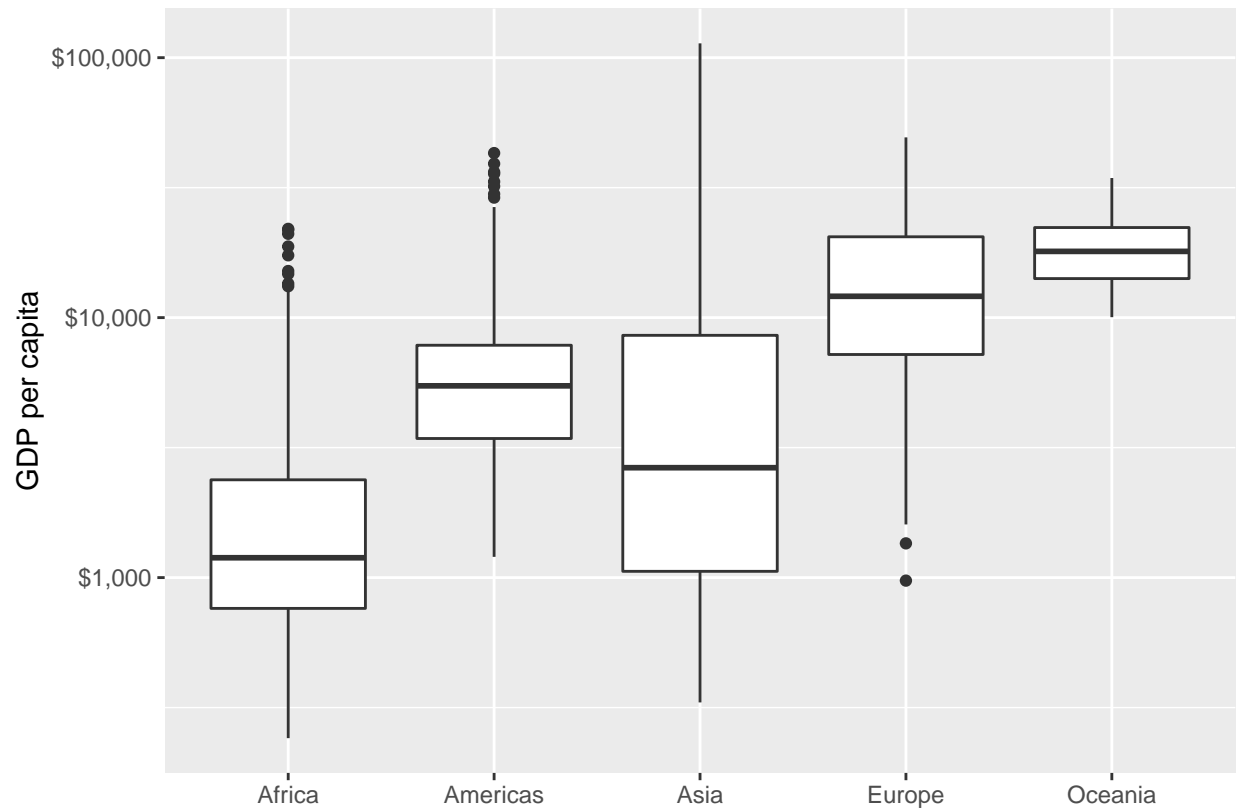
continent	mean_gdp	sd_gdp	rsd_gdp	mean_pop	sd_pop	rsd_pop
Africa	2193.75	2827.93	128.91	9916003	15490923	156.22
Americas	7136.11	6396.76	89.64	24504795	50979430	208.04
Asia	7902.15	14045.37	177.74	77038722	206885205	268.55
Europe	14469.48	9355.21	64.65	17169765	20519438	119.51
Oceania	18621.61	6358.98	34.15	8874672	6506342	73.31

```

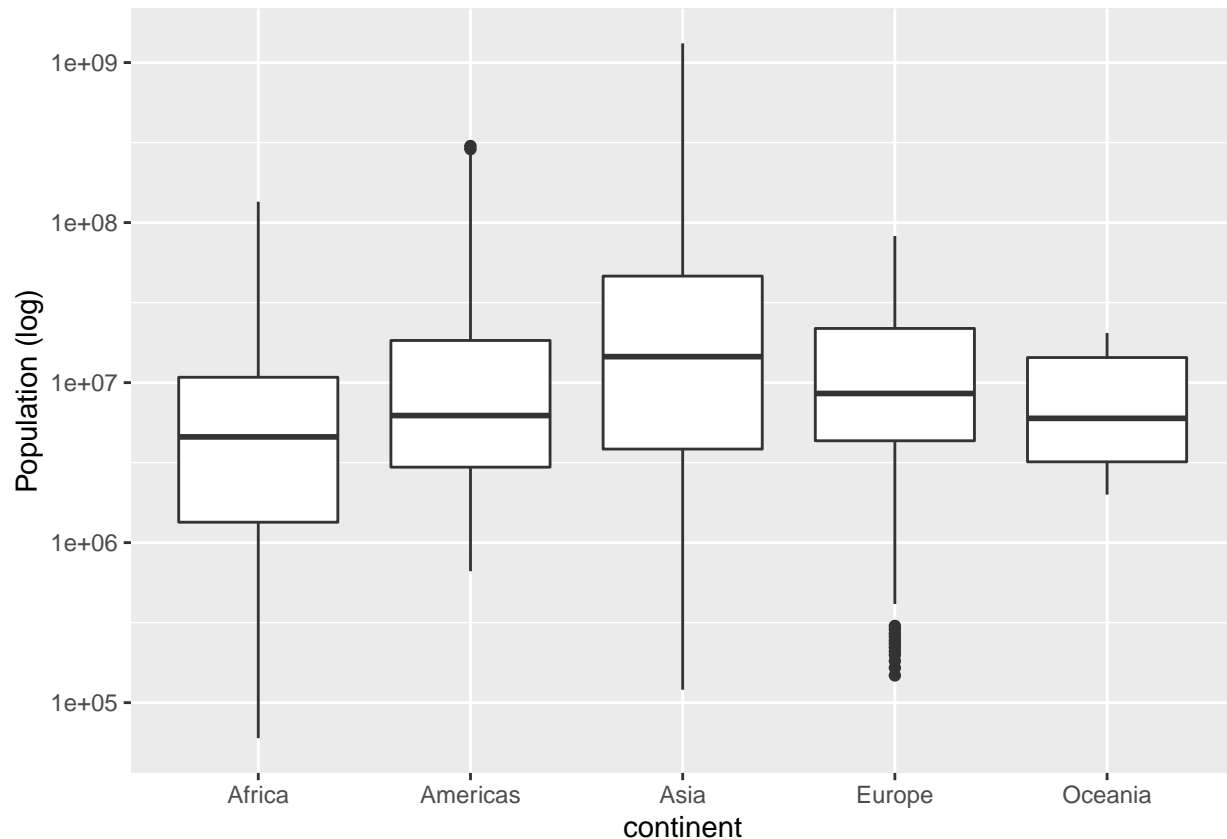
# yuck. that's a lot of numbers

# try plotting it
# put GDP on a log scale again
gapminder %>%
  ggplot(aes(x = continent, y = gdpPercap)) +
  geom_boxplot() +
  labs(y = "GDP per capita",
       x = "") +
  scale_y_log10("GDP per capita", labels = scales::dollar_format())

```



```
# plot population also, because I think they're related
gapminder %>%
  ggplot(aes(x = continent, y = pop)) +
  geom_boxplot() +
  scale_y_log10("Population (log)")
```



### Exercise 3: Explore various plot types

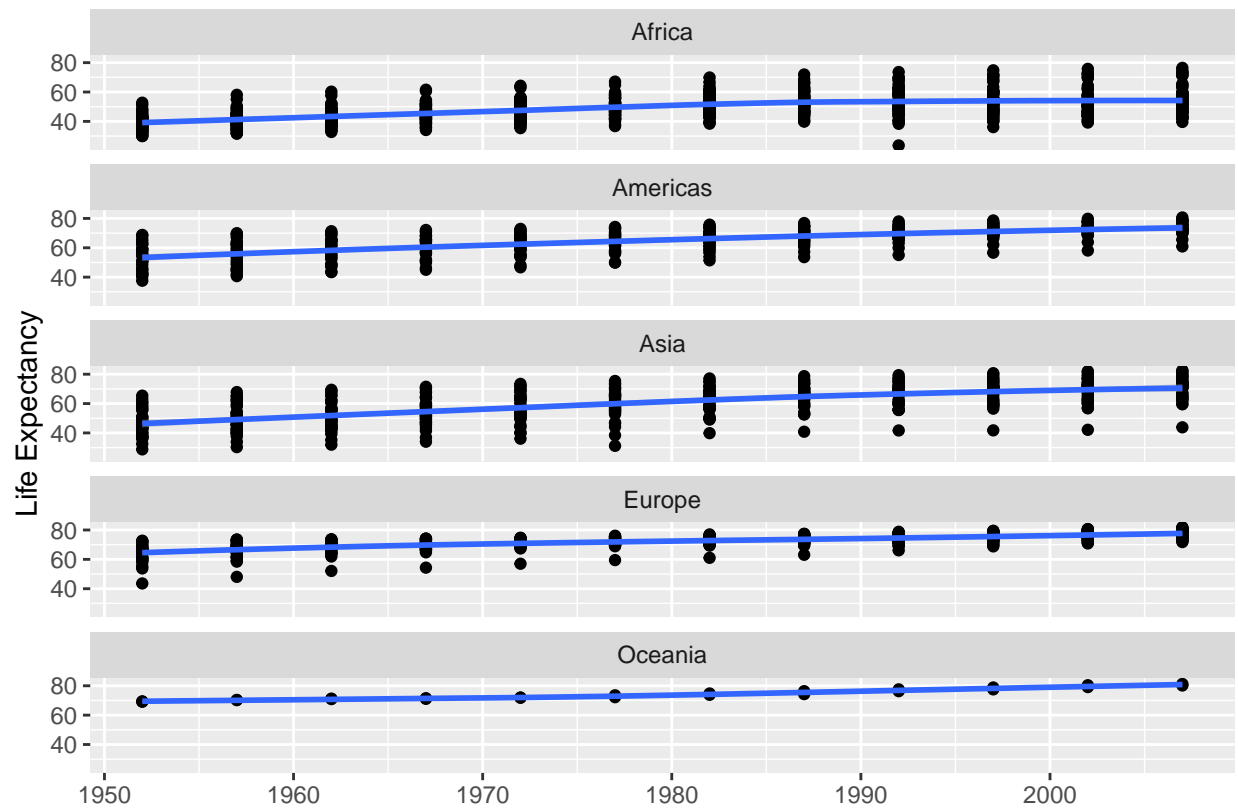
Make two plots that have some value to them. That is, plots that someone might actually consider making for an analysis. Just don't make the same plots we made in class – feel free to use a data set from the `datasets` R package if you wish.

- A scatterplot of two quantitative variables.
- One other plot besides a scatterplot.

Note: You don't have to use all the data in every plot! It's fine to filter down to one country or a small handful of countries.

```
# 2 scatter plots
# these are similar facet plots like I plotted in Exercise 2 -- because I like it

# life expectancy per capita over time
gapminder %>%
  ggplot(aes(x = year, y = lifeExp)) +
  geom_point() +
  geom_smooth(method = "loess") +
  facet_wrap(~ continent, ncol = 1) +
  labs(x = "", y = "Life Expectancy")
```



*# GDP per capita over time*

gapminder %>%

ggplot(aes(x = year, y = gdpPercap)) +

geom\_point() +

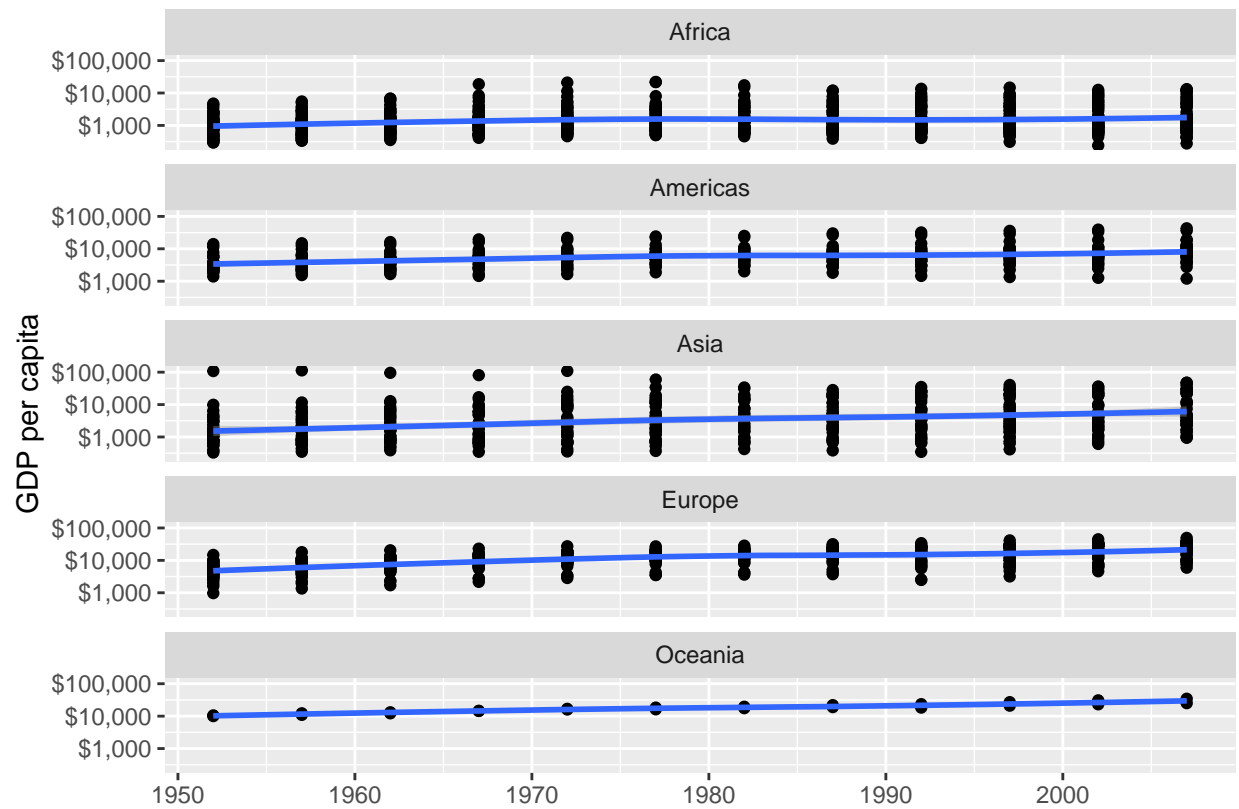
geom\_smooth(method = "loess") +

facet\_wrap(~ continent, ncol = 1) +

scale\_y\_log10("GDP per capita", labels = scales::dollar\_format()) +

labs(x = "")

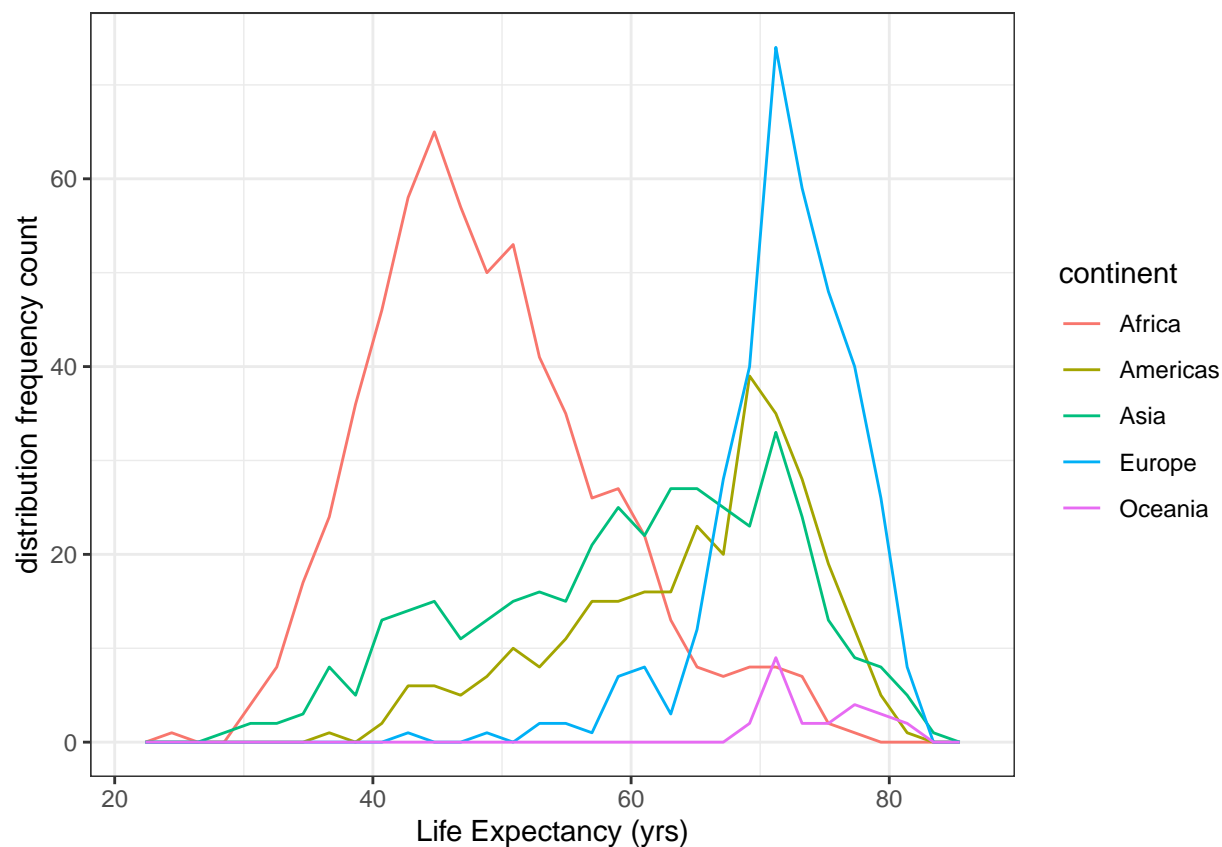




One other plot (not a scatter plot)

```
# histograms show distribution of a quantitative variable
# 'geom_freqpoly()'
# shows distribution like a histogram but with lines instead of shaded bars
gapminder %>%
  ggplot(aes(x = lifeExp, color = continent)) +
  theme_bw() +
  labs(x = "Life Expectancy (yrs)", y = "distribution frequency count") +
  geom_freqpoly()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



## Recycling (Optional)

Rubrics:

reasoning

How well did you reason through the scenario?

For people who want to take things further.

Evaluate this code and describe the result. Presumably the analyst's intent was to get the data for Rwanda and Afghanistan. Did they succeed? Why or why not? If not, what is the correct way to do this?

```
# provided code
filter(gapminder, country == c("Rwanda", "Afghanistan")) %>%
  kable(digits = 2, caption = "original code")
```

Table 9: original code

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1957	30.33	9240934	820.85
Afghanistan	Asia	1967	34.02	11537966	836.20
Afghanistan	Asia	1977	38.44	14880372	786.11
Afghanistan	Asia	1987	40.82	13867957	852.40

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1997	41.76	22227415	635.34
Afghanistan	Asia	2007	43.83	31889923	974.58
Rwanda	Africa	1952	40.00	2534927	493.32
Rwanda	Africa	1962	43.00	3051242	597.47
Rwanda	Africa	1972	44.60	3992121	590.58
Rwanda	Africa	1982	46.22	5507565	881.57
Rwanda	Africa	1992	23.60	7290203	737.07
Rwanda	Africa	2002	43.41	7852401	785.65

*# I added the kable call for easier comparison between this and my code below*

It does not appear that this method of filtering delivered the desired results. It looks like data from Rwanda and Afghanistan for overlapping years were omitted, resulting in each country having data provided every decade: Afghanistan had data for every ten years ending in 7 (e.g. 1957, 1967, 1977, etc.) while Rwanda had data for every ten years ending in 2 (e.g. 1952, 1962, 1972, etc.). I think the problem is associated with trying to filter country based on a vector of names. When the country filter is explicitly defined as Rwanda OR Afghanistan, the result is a tibble twice as long as the original output. This method results in data for both countries on overlapping years.

```
# try some other options to see if the results agree
gapminder %>%
  filter(country == "Rwanda" | country == "Afghanistan") %>%
  kable(digits = 2, caption = "modified code")
```

Table 10: modified code

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1952	28.80	8425333	779.45
Afghanistan	Asia	1957	30.33	9240934	820.85
Afghanistan	Asia	1962	32.00	10267083	853.10
Afghanistan	Asia	1967	34.02	11537966	836.20
Afghanistan	Asia	1972	36.09	13079460	739.98
Afghanistan	Asia	1977	38.44	14880372	786.11
Afghanistan	Asia	1982	39.85	12881816	978.01
Afghanistan	Asia	1987	40.82	13867957	852.40
Afghanistan	Asia	1992	41.67	16317921	649.34
Afghanistan	Asia	1997	41.76	22227415	635.34
Afghanistan	Asia	2002	42.13	25268405	726.73
Afghanistan	Asia	2007	43.83	31889923	974.58
Rwanda	Africa	1952	40.00	2534927	493.32
Rwanda	Africa	1957	41.50	2822082	540.29
Rwanda	Africa	1962	43.00	3051242	597.47
Rwanda	Africa	1967	44.10	3451079	510.96
Rwanda	Africa	1972	44.60	3992121	590.58
Rwanda	Africa	1977	45.00	4657072	670.08
Rwanda	Africa	1982	46.22	5507565	881.57
Rwanda	Africa	1987	44.02	6349365	847.99
Rwanda	Africa	1992	23.60	7290203	737.07
Rwanda	Africa	1997	36.09	7212583	589.94
Rwanda	Africa	2002	43.41	7852401	785.65
Rwanda	Africa	2007	46.24	8860588	863.09

## Tibble display (Optional)

Present numerical tables in a more attractive form using `knitr::kable()` for small tibbles (say, up to 10 rows), and `DT::datatable()` for larger tibbles.

*I believe I incorporated this into my code already :smiley:*