HW-02-dplyr

```
library(gapminder)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Exercise 1

Exercise 1.1

```
gapminder %>%
     filter(country %in% c("Japan", "Canada", "Kenya")) %>%
     filter(year %in% c("1972", "1977"))
```

```
## # A tibble: 6 x 6
##   country continent  year lifeExp       pop gdpPercap
##   <fct>   <fct>     <int>   <dbl>     <int>     <dbl>
## 1 Canada  Americas   1972    72.9  22284500    18971.
## 2 Canada  Americas   1977    74.2  23796400    22091.
## 3 Japan   Asia       1972    73.4 107188273    14779.
## 4 Japan   Asia       1977    75.4 113872473    16610.
## 5 Kenya   Africa     1972    53.6  12044785     1222.
## 6 Kenya   Africa     1977    56.2  14500404     1268.
```

Exercise 1.2

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

```
gapminder %>%
     filter(country %in% c("Japan", "Canada", "Kenya")) %>%
     filter(year %in% c("1972", "1977")) %>%
     select(country, gdpPercap)
```

```
## # A tibble: 6 x 2
##   country gdpPercap
##   <fct>       <dbl>
## 1 Canada     18971.
## 2 Canada     22091.
## 3 Japan      14779.
## 4 Japan      16610.
## 5 Kenya       1222.
```

```
## 6 Kenya       1268.
```

Exercise 1.3

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.

```
gapminder %>%
  group_by(country) %>%
  arrange(year) %>%
  mutate(inc_lifeExp = diff(lifeExp)) %>%
  drop_na() %>%
  filter(inc_life_Exp == min(inc_life_Exp))
```

```
## Error: Column `inc_lifeExp` must be length 12 (the group size) or one, not 11
```

Exercise 1.4

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

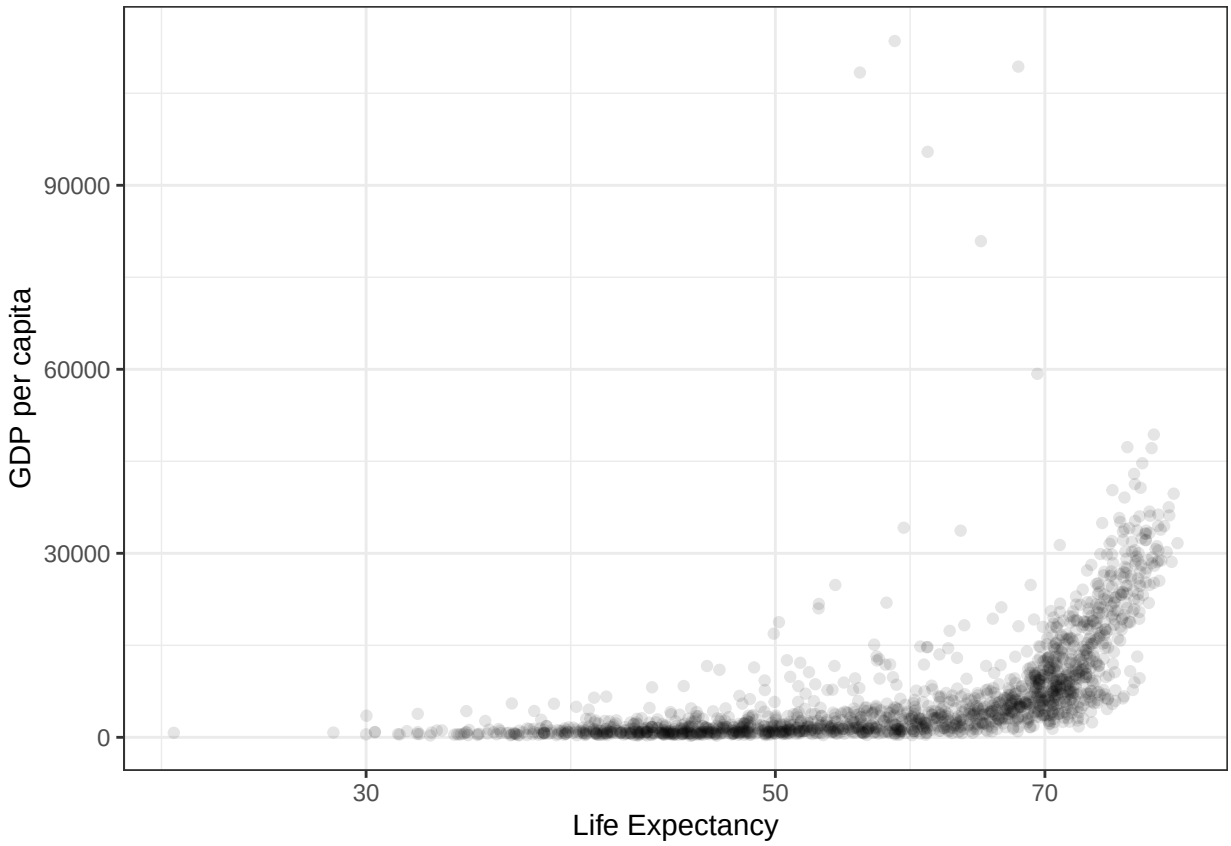
```
gapminder %>%
  filter(country) %>%
  max(gdpPercap)
```

```
## Error: Argument 2 filter condition does not evaluate to a logical vector
```

Exercise 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.

```
ggplot(gapminder, aes(lifeExp, gdpPercap)) +
  geom_point(alpha = 0.1) +
  scale_x_log10("Life Expectancy") +
  theme_bw() +
  ylab("GDP per capita")
```



Exercise 2: Explore individual variables with 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? Categorical variable: Continent Possible values: Africa, Americas, Asia, Europe, Oceania.

Quantitative variable: Life Expectancy: Range 23.6 - 82.6

2. What values are typical? What's the spread? What's the distribution? Etc., tailored to the variable at hand. Most common continent is African by far. Least common is Oceania by far. Average is Europe. Mean life expectancy is 59.47.

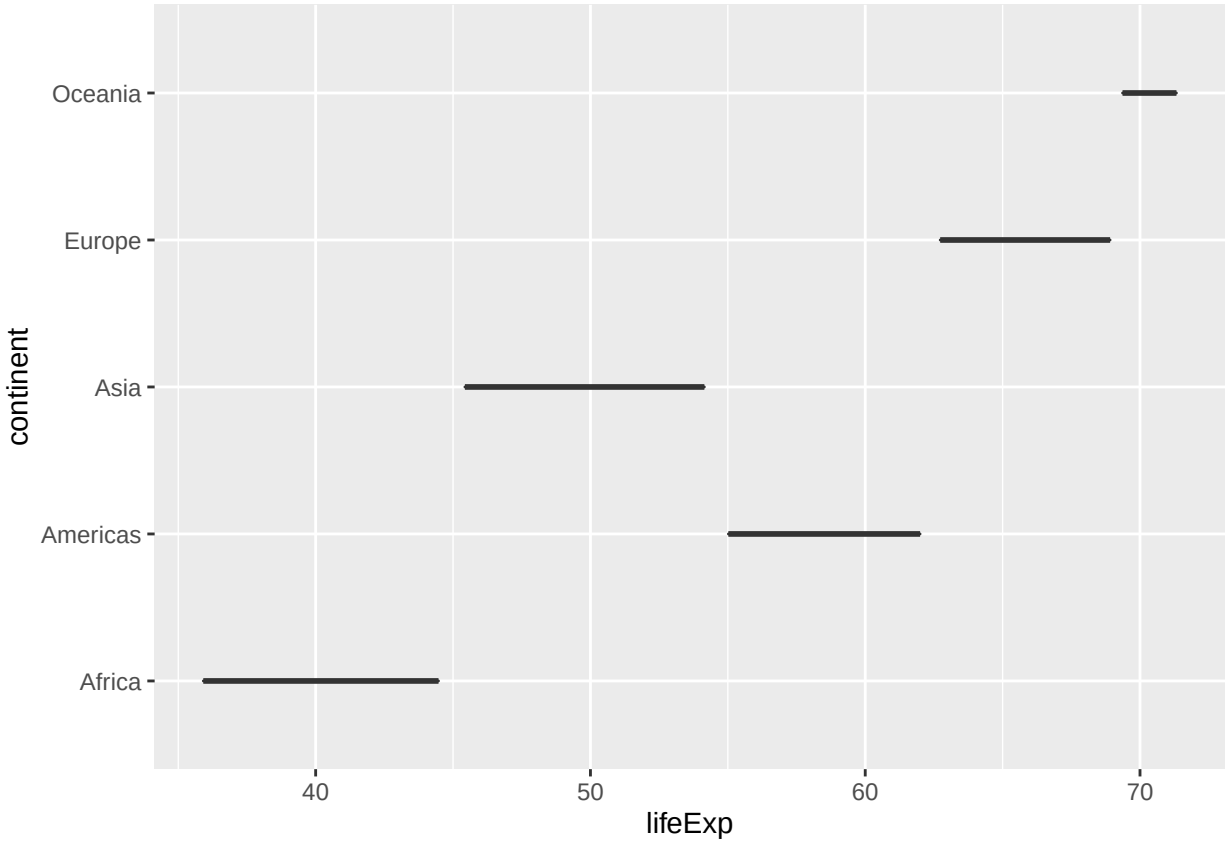
3. Feel free to use summary stats, tables, figures.

```
summary(gapminder)
```

```
##         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
##
```

```
ggplot(gapminder, aes(lifeExp, continent)) +
  geom_boxplot()
```

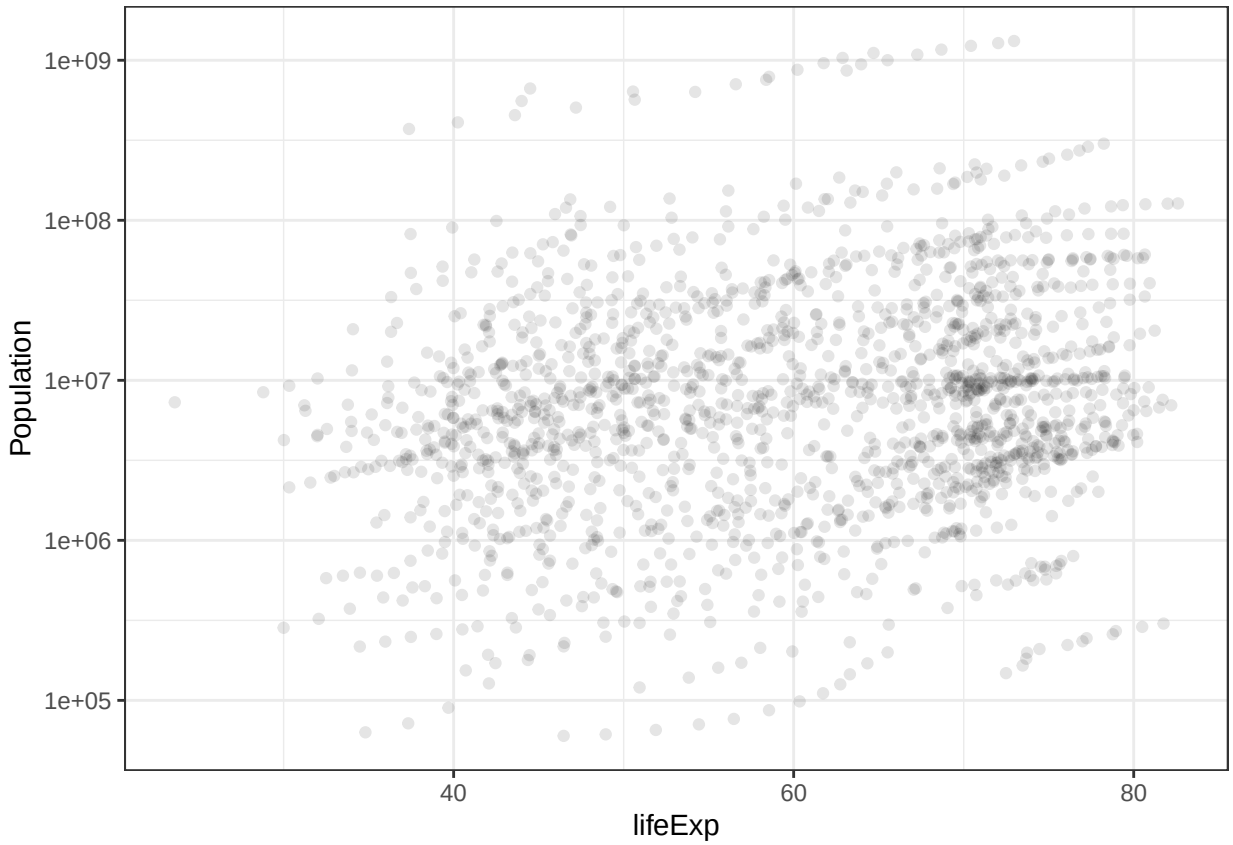


Exercise 3: Explore various plot types (30%)

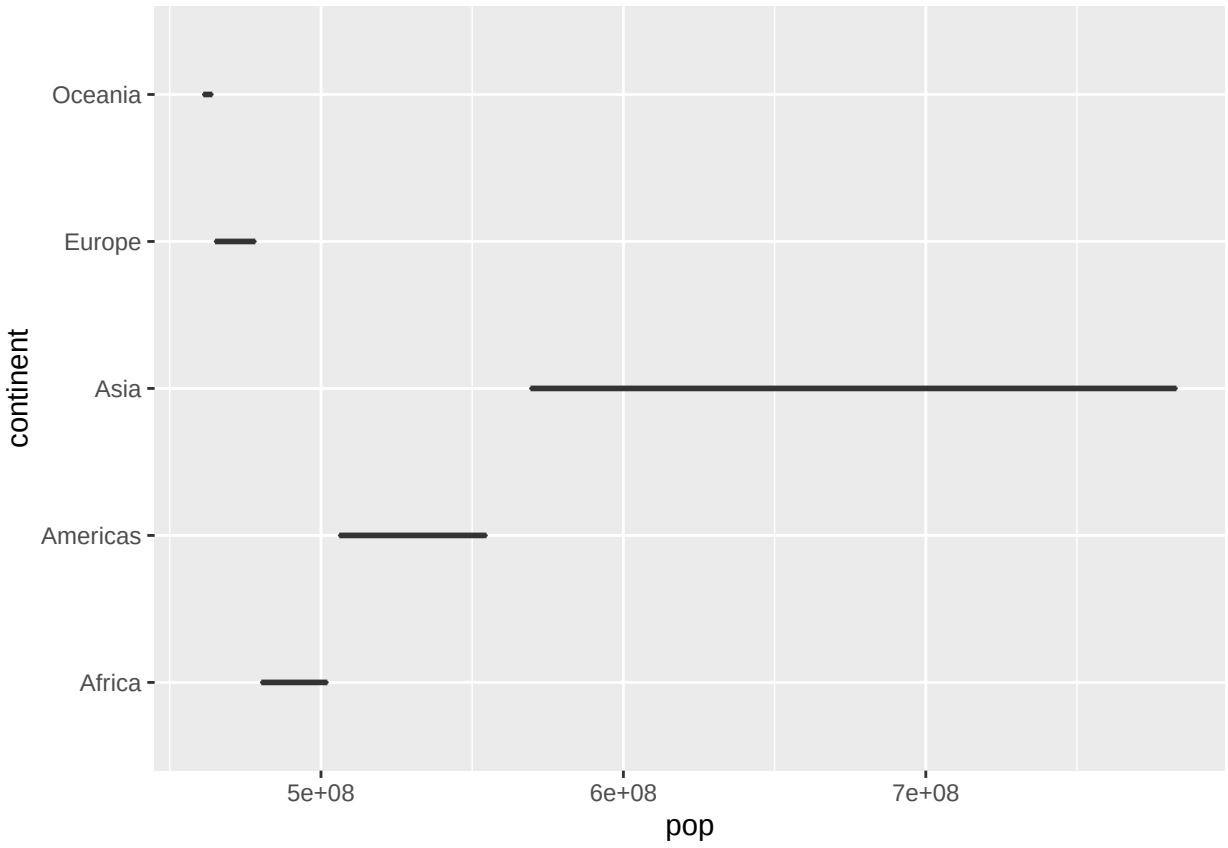
Make two plots that have some value to them. That is, plots that someone might actually consider making for an analysis. Just don't make the same plots we made in class – feel free to use a data set from the datasets R package if you wish.

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

```
ggplot(gapminder, aes(lifeExp, pop)) +
  geom_point(alpha = 0.1) +
  theme_bw() +
  scale_y_log10("Population")
```




```
ggplot(gapminder, aes(pop, continent)) +
  geom_boxplot()
```


Oceania
Europe
Asia
Americas
Africa
continent
5e+08
6e+08
7e+08
pop