hw02

JMF

27/09/2019

1.1 and 1.2

Use filter() to subset the gapminder data to three countries of your choice in the 1970's. Use the pipe operator %>% to select "country" and "gdpPercap" from your filtered dataset in 1.1.

```
subset <- filter(gapminder, country %in%
                   c("Canada", "Mexico", "United States"),
                 year > 1969,  year < 1981)
subset %>% select(country,gdpPercap)
```

```
## # A tibble: 6 x 2
##   country       gdpPercap
##   <fct>             <dbl>
## 1 Canada           18971.
## 2 Canada           22091.
## 3 Mexico            6809.
## 4 Mexico            7675.
## 5 United States    21806.
## 6 United States    24073.
```

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.

```
le_drop <- select(gapminder, country, year, continent, lifeExp) %>%
  group_by(country) %>%
  mutate(le_delta = lifeExp - lag(lifeExp))

  filter(le_drop, le_delta<0)
```

```
## # A tibble: 102 x 5
## # Groups:   country [52]
##    country   year continent lifeExp le_delta
##    <fct>    <int> <fct>       <dbl>    <dbl>
##  1 Albania   1992 Europe       71.6   -0.419
##  2 Angola    1987 Africa       39.9   -0.036
##  3 Benin     2002 Africa       54.4   -0.371
##  4 Botswana  1992 Africa       62.7   -0.877
##  5 Botswana  1997 Africa       52.6  -10.2
##  6 Botswana  2002 Africa       46.6   -5.92
##  7 Bulgaria  1977 Europe       70.8   -0.09
##  8 Bulgaria  1992 Europe       71.2   -0.15
##  9 Bulgaria  1997 Europe       70.3   -0.87
## 10 Burundi   1992 Africa       44.7   -3.48
## # ... with 92 more rows
```

1.4a

Max GDP per capita experienced by each country. Filter gapminder so that it shows the max GDP per capita experienced by each country.

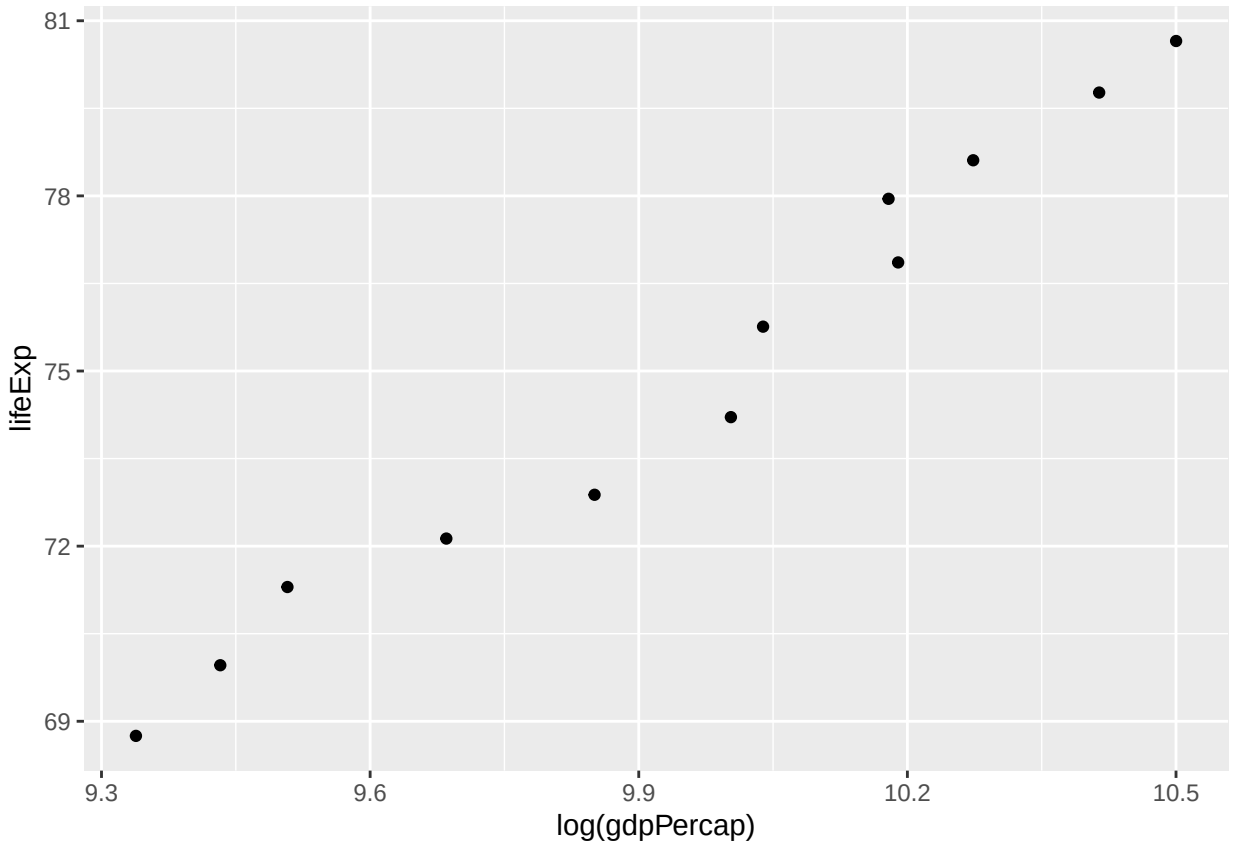
```
  gapminder %>%
  select(country, year, continent, gdpPercap) %>%
  arrange(country) %>%
  group_by(country) %>%
  top_n(1, wt = desc(gdpPercap)) ## gets the max
```

```
## # A tibble: 142 x 4
## # Groups:   country [142]
##    country      year continent gdpPercap
##    <fct>       <int> <fct>         <dbl>
##  1 Afghanistan  1997 Asia           635.
##  2 Albania      1952 Europe        1601.
##  3 Algeria      1952 Africa        2449.
##  4 Angola       1997 Africa        2277.
##  5 Argentina    1952 Americas      5911.
##  6 Australia    1952 Oceania      10040.
##  7 Austria      1952 Europe        6137.
##  8 Bahrain      1952 Asia          9867.
##  9 Bangladesh   1972 Asia           630.
## 10 Belgium      1952 Europe        8343.
## # ... with 132 more rows
```

1.5

Scatterplot of Canada's life expectancy vs. GDP per capita using ggplot2

```
filter(gapminder, country %in% c("Canada")) %>%
ggplot(aes(x=log(gdpPercap),y=lifeExp)) + geom_point()
```



2

Explore individual variables with dplyr

```
#summary table for Life Exp.

summary_lifeexp <-
  list("Life Expentancy" =
         list("min" = ~ min(gapminder$lifeExp),
              "max" = ~ max(gapminder$lifeExp),
              "mean" = ~ mean(gapminder$lifeExp),
              "sd" = ~ sd(gapminder$lifeExp)),
       "Years" =
         list("min" = ~ min(gapminder$year),
              "max" = ~ max(gapminder$year)
              ))


summary_table(gapminder, summary_lifeexp)
```

```
##
##
## |                   |gapminder (N = 1,704) |
## |:------------------|:---------------------|
## |**Life Expentancy** |            |
## |   min    |23.599                |
```

```
## |   max    |82.603               |
## |   mean   |59.4744393661972     |
## |   sd     |12.9171074152412     |
## |**Years**           |           |
## |   min    |1952                 |
## |   max    |2007                 |
```

2a

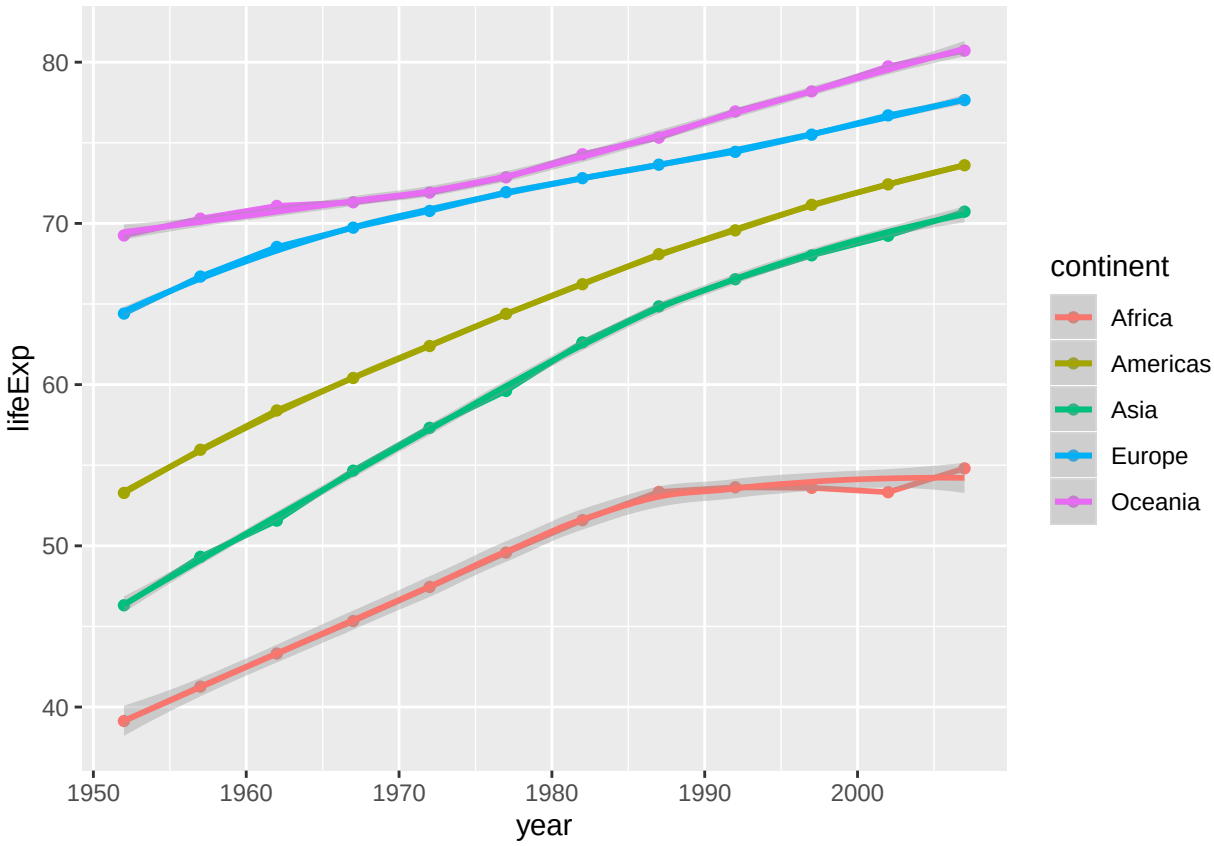
Grouped by Continent

```
summary_table(dplyr::group_by(gapminder, continent), summary_lifeexp)
```

```
##
##
## |                   |continent: Africa (N = 624) |continent: Americas (N = 300) |continent: Asia (N
## |:------------------|:---------------------------|:-----------------------------|:-----------------
## |**Life Expentancy** |                 |                    |  
## |   min    |23.599                      |23.599                        |23.599
## |   max    |82.603                      |82.603                        |82.603
## |   mean   |59.47444                    |59.47444                      |59.47444
## |   sd     |12.91711                    |12.91711                      |12.91711
## |**Years**           |                  |                    |  
## |   min    |1952                        |1952                          |1952
## |   max    |2007                        |2007                          |2007
```

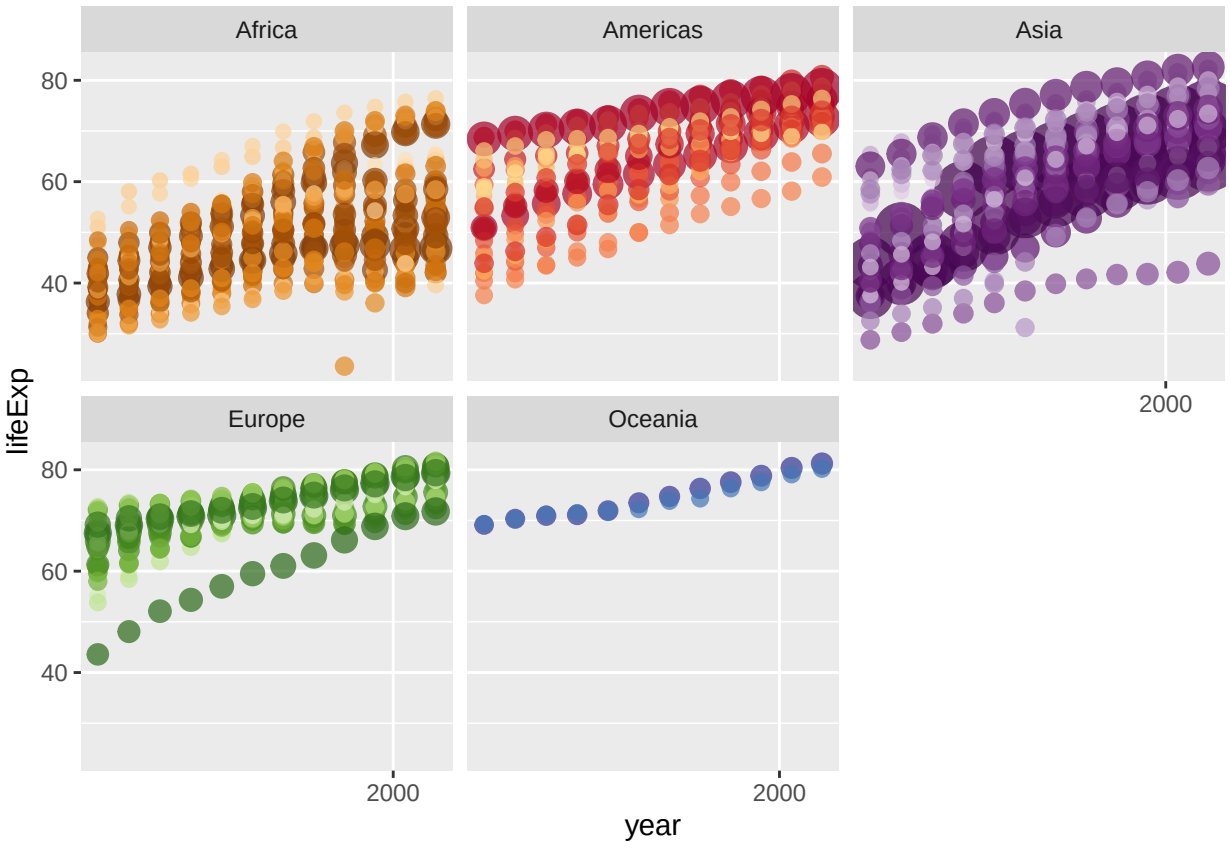
3

Average Life expectancy in time by continent



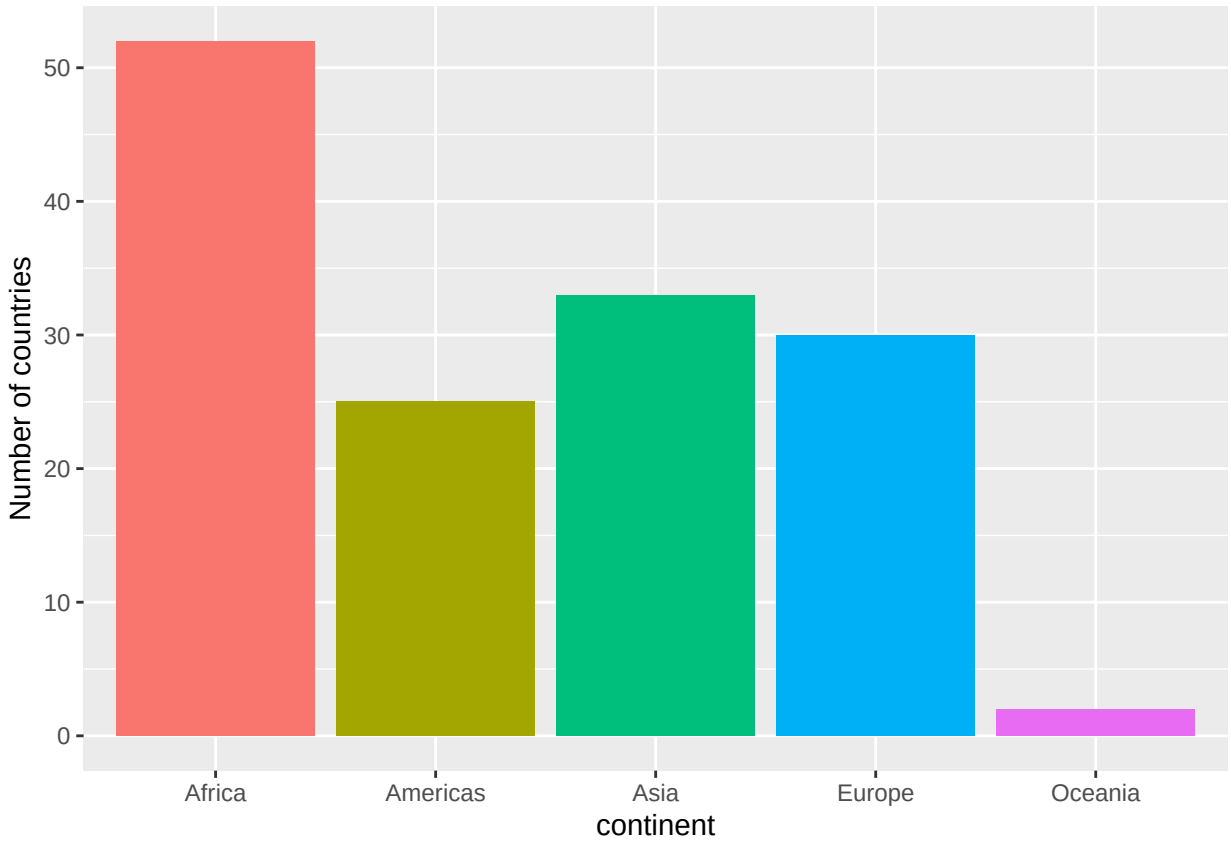
3

Life expectancy of each country in time by continent



3

Frequency of contries by continent



4

This is the correct code. Here %in% instead of == makes the filter to take all elements in country that match the countries Rwanda and Afghanistan

```
filter(gapminder, country %in% c("Rwanda", "Afghanistan"))
```

```
## # A tibble: 24 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # ... with 14 more rows
```